

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2004, 08:55:46 ; Search time 53 Seconds
(without alignments)
421.156 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456
Sequence: 1 MCGPRRVSACGFADAHWT.....SAGLTVRDPLQLGCMGRG 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	79	ABD08035	ABD08035 Human ERC
2	82.5	18.1	168	ABR56758	ABR56758 Human sec
3	79.5	17.4	151	ABG22219	ABG22219 Novel hum
4	78	17.1	11096	AAE10129	AAE10129 Streptomy
5	77.5	17.0	426	ABG16164	ABG16164 Novel hum
6	77	16.9	215	ADA54191	ADA54191 Human pro
7	77	16.9	235	ABP95644	ABP95644 Human GPC
8	76.5	16.8	206	ABG14023	ABG14023 Novel hum
9	75.5	16.6	443	ABG03615	ABG03615 Novel hum
10	75.5	16.6	1063	ADC86843	ADC86843 Human GPC
11	75.5	16.6	1356	ADC86885	ADC86885 Human GPC
12	74.5	16.3	104	ABM65289	ABM65289 Propionib
13	74.5	16.3	156	ABP69309	ABP69309 Human pol
14	74.5	16.3	180	ABG14843	ABG14843 Novel hum
15	74.5	16.3	180	ABG12827	ABG12827 Novel hum
16	74.5	16.3	329	AAO10468	AAO10468 Human pol
17	74	16.2	393	ABG97894	ABG97894 Human sec
18	73.5	16.1	1180	ADC30896	ADC30896 Human nov
19	72.5	15.9	172	AAU65951	AAU65951 Propionib
20	72.5	15.9	172	ABM62270	ABM62270 Propionib
21	72.5	15.9	372	ADC07760	ADC07760 Rice prot
22	72	15.8	371	ABG03350	ABG03350 Novel hum
23	72	15.8	384	ABG06095	ABG06095 Novel hum
24	72	15.8	514	AAW25917	AAW25917 Human pro
25	71.5	15.7	421	ABG19501	ABG19501 Novel hum

26	71	15.6	93	4	AAU45659	AAU45659 Propionib
27	71	15.6	93	6	ABM42178	ABM42178 Propionib
28	71	15.6	192	4	ABG26053	ABG26053 Novel hum
29	71	15.6	253	4	AAE03237	AAE03237 Human gen
30	71	15.6	253	5	ABG64383	ABG64383 Human alb
31	71	15.6	359	4	ABM60902	ABM60902 Drosophil
32	70.5	15.5	638	2	AAU04992	AAU04992 Mycobacte
33	70	15.4	532	2	AAW40114	AAW40114 Human alp
34	70	15.4	1214	4	ABG61851	ABG61851 Drosophil
35	70	15.4	1669	4	AAW40863	AAW40863 Human pol
36	70	15.4	1669	5	ABB90760	ABB90760 Human tum
37	70	15.4	1669	6	ABU54467	ABU54467 Human tum
38	70	15.4	1672	4	AAW39077	AAW39077 Human pol
39	69.5	15.2	170	2	AAZ29195	AAZ29195 Amino aci
40	69.5	15.2	552	4	AAW75591	AAW75591 Human aci
41	69	15.1	101	3	AAW93886	AAW93886 Amino aci
42	69	15.1	19938	6	ABP76681	ABP76681 Streptomy
43	68.5	15.0	207	5	ABJ10911	ABJ10911 Human sec
44	68.5	15.0	207	7	ADB64597	ADB64597 Human pro
45	68.5	15.0	591	4	ABG03618	ABG03618 Novel hum

ALIGNMENTS

RESULT 1
ABD08035
ID ABD08035 standard; protein; 79 AA.
XX
AC ABD08035;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human ERCoA3 protein.
XX
KW Estrogen Receptor Coregulator 3; ERCoA3; tamoxifen; estrogen; cancer;
KW osteoporosis; cytostatic; osteopathic; human; receptor.
OS Homo sapiens.
XX
FN W0200228352-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US031271.
XX
PR 05-OCT-2000; 2000US-0238190P.
XX (UYCA-) UNIV CASE WESTERN RESERVE.
PA Montano M, Sutton A;
XX
DR WPI; 2002-454492/48.
DR N-PSDB; ABL06066.
XX
PT New polypeptide, that is a functional equivalent of ERCoA3 (Estrogen Receptor Coregulator 3), is useful in inhibiting or reducing tamoxifen or estrogen-induced proliferation of cancer cells and in treating osteoporosis.
XX
PS Claim 1; Fig 2; 39pp; English.
XX
CC The invention relates to a ERCoA3 (Estrogen Receptor Coregulator 3) protein and encoding polynucleotides. ERCoA3 can be used to inhibit or reduce tamoxifen or estrogen induced proliferation of cancer cells, by reduced activity of ERCoA3, and for detecting cancer cells that are tamoxifen resistant, or to treat osteoporosis, by increasing levels of ERCoA3 in cells. The encoding polynucleotide can be used to inhibit ERCoA3 translation of a mRNA encoding ERCoA3. ERCoA3 acts as a coregulator protein and can bind to the estrogen receptor to activate a molecular or cellular response in the cell. The present sequence represents the human ERCoA3 protein

Claim 1; Page 215; 286pp; English.

reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polynucleotide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published](http://wipo.int/pub/published) pct sequences

Sequence 151 AA;

Query Match	17.4%;	Score 79.5;	DB 4;	Length 151;
Best Local Similarity	37.0%;	Pred. No. 0.83;		

QY 27 GEGQEGGIGPEQA-----SPTPDCASRWPRASRWPSAGLTVR 66

Db 55 GAGREHGMGPCGQAQPLSKLSPGPKC--WVEQMHRWTWHTQLSLK 97

RESULT 4

AE101029
ID AE101029 standard; protein; 11096 AA.

XX
AC AAE10129;

AA	DT	29-NOV-2001 (first entry)
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

DE Streptomyces noursei nystatin gene, NysC.

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 KW antifungal; antibiotic; PKS type I.
 KW

Streptomyces noursei.

Key	Location/Qualifiers
XX	
FT	35..455
FT	/label= KS3 domain
FT	/note= "ketosynthase (KS) domain"
FT	546..858
FT	/label= AT3 domain
FT	/note= "acyltransferase (AT) domain"
FT	872..1073
FT	/label= DH3 domain
FT	/note= "dehydratase (DH) domain"
FT	1381..1528
FT	/label= KR3 domain
FT	/note= "ketoreductase (KR) domain"
FT	1562..1735
FT	/label= ACP3 domain
FT	/note= "Acyl carrier protein (ACP) domain"
FT	1757..2180
FT	/label= KS4 domain
FT	/note= "ketosynthase (KS) domain"
FT	2291..2603
FT	/label= AT4 domain
FT	/note= "acyltransferase (AT) domain"
FT	2617..2818
FT	/label= BH4 domain
FT	/note= "dehydratase (DH) domain"
FT	3124..3371
FT	/label= KR4 domain
FT	

XX 08-FEB-2000; 2000GB-00002840.
 PR 10-APR-2000; 2000GB-00008786.
 PR 14-APR-2000; 2000GB-00009387.
 XX
 PA (UNNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
 PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
 PA (ALPH-) ALPHARMA AS.
 PA (SINV-) SINVENT AS.
 PA (DZIE/) DZIEGLEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAEVRIK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 PA (VALL/) VALLA S.
 XX
 PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX
 DR WPI; 2001-557614/62.
 DR N-PSDB; AAD17184.
 XX
 PT New nystatin polyketide synthase polynucleotides and polypeptides, useful
 PT as antibiotics and antifungals.
 XX
 PS Claim 15; Page 170-176; 266pp; English.
 CC The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrobide antibiotic nystatin. The
 CC nystatin PKS is useful as antifungal antibiotics. The present sequence is
 CC a PKS type I encoding Streptomyces noursei nystatin gene, Nysc
 XX
 SQ Sequence 11096 AA;
 Query Match 17.1%; Score 78; DB 4; Length 11096;
 Best Local Similarity 37.2%; Pred. No. 1.3e+02;
 Matches 32; Conservative 7; Mismatches 33; Indels 14; Gaps 6;
 Qy 4 RPRVSAAGCGPADAHWTGL---WTGLGEGGEGGIGPEGQASPT-PDCASR-WPRS--ASR 56
 Db 836 RPELSAVTGLARHVRGVTVRWAGLFD---GTGARRADLTPTYPFHQRFWPTAARAAQ 891
 Qy 57 WPSAGLTVDRDPOLG---ELCMGRG 79
 Db 892 DVTAAAGLGAADHPLIGATVELADGAG 917
 RESULT 5
 ABG16164
 ID ABG16164 standard; protein; 426 AA.
 XX
 AC ABG16164;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #16155.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX

PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS80351.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 46523; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ASG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: the sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 426 AA;
 Query Match 17.0%; Score 77.5; DB 4; Length 426;
 Best Local Similarity 37.0%; Pred. No. 4.2;
 Matches 17; Conservative 2; Mismatches 18; Indels 9; Gaps 1;
 Qy 36 PRGQASPTPDCASRWPF-----RSASRWPGAGLTVDRDPOLG 72
 Db 194 PPHLPSPDGLQGPSRIVGAVSSEGEWPGASLQVRCRHLG 239
 RESULT 6
 ADA54191
 ID ADA54191 standard; protein; 215 AA.
 XX
 AC ADA54191;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human protein, SEQ ID 1759.
 XX
 KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
 KW inflammatory disease; osteoporosis; neurological disease.
 XX
 OS Homo sapiens.
 XX
 PN EP1293569-A2.
 XX
 PD 19-MAR-2003.
 XX
 PF 21-MAR-2002; 2002EP-00006586.
 XX
 PR 14-SEP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 PA (HELI-) HELIX RES INST.

(REAS-) RES ASSOC BIOTECHNOLOGY.

PA Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX
 DR WPI; 2003-395539/38.
 DR N-PSDB; ADA52552.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX
 PS Claim 14; SEQ ID NO 1759; 205pp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 XX
 SQ Sequence 215 AA;

Query Match 16.9%; Score 77; DB 6; Length 215;
 Best Local Similarity 31.9%; Pred. No. 2.3;
 Matches 23; Conservative 3; Mismatches 28; Indels 18; Gaps 4;

QY 7 RVSAAGCGFADAHWTGL-----WTGLGEGGEGGIGPEGOASPTPDC-ASRWPRSRASRW 57
 DB 64 RCGARC-----WEGSSLAQLQPPWTPSGPSLTAAVGPQVCA---SCWRSWPRSCPRW 114
 QY 58 FWSAGLTVDRRP 69
 DB 115 PPSSTFAAARSSP 126

RESULT 7
 ABP95644
 ID ABP95644 standard; protein; 235 AA.
 AC ABP95644;
 XX
 DT 06-MAR-2003 (first entry)
 XX
 DE Human GPCR polypeptide SEQ ID NO 98.
 XX
 KW Human; GPCR; G protein coupled receptor; signal transduction; olfactory;
 KW drug development; gustatory; taste; fragrance; receptor.
 XX
 OS Homo sapiens.
 XX
 PN WO200216548-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 30-JUL-2001; 2001WO-IB001446.
 XX
 PR 04-AUG-2000; 2000JP-00237818.
 PR 13-FEB-2001; 2001JP-00034434.
 XX
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 XX
 PI Haga T, Takeda S, Mitaku S;
 XX
 DR WPI; 2002-304118/34.
 DR N-PSDB; AB242918.
 XX
 PT Database global search for G protein-coupled receptors, proteins and
 PT encoded genes for studying in vivo signal transduction mechanism and
 PT identifying targets for drug development.
 XX
 PS Claim 10; SEQ ID NO 98; 97pp + Sequence Listing; Japanese.
 XX

CC The invention relates to a method for screening G protein-coupled
 CC receptor (GPCR) genes (AB242870-AB243216) and/or GPCR proteins (ABP95596-
 CC ABP95942) by extracting open-reading frames containing 6-8 transmembrane
 CC domains with 250-1000 amino acid residues to give a gene homologous with
 CC a known GPCR gene. The receptor proteins and encoded genes are useful for
 CC studying in vivo signal transduction mechanism and identifying targets
 CC for drug development e.g. based on olfactory and gustatory receptors in
 CC form of agonists and antagonists by screening intrinsic and extrinsic
 CC ligands as bitter taste inhibitors, taste enhancers and fragrance
 CC improvers. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 235 AA;

Query Match 16.9%; Score 77; DB 5; Length 235;
 Best Local Similarity 37.2%; Pred. No. 2.5;
 Matches 16; Conservative 3; Mismatches 18; Indels 6; Gaps 1;

QY 19 WTGLWTGLGEGGIGPEGOASPTPDCASRWPRSRASRWPSA 61
 DB 11 WLGLWVGL-----GLRPTFRVCSPLCGPLWPRGASLCVWGS 47

RESULT 8
 ABG14023
 ID ABG14023 standard; protein; 206 AA.
 AC ABG14023;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #14014.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS78210.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 44382; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 206 AA;

Query Match 16.8%; Score 76.5; DB 4; Length 206;
 Best Local Similarity 33.3%; Pred. No. 2.5;
 Matches 25; Conservative 3; Mismatches 28; Indels 19; Gaps 3;
 QY 5 PRRVSAGC-GFADAHWTGLWTG-----LGGQGG--GIGPQGQASPTPD 45
 DB 91 PREGXGSRVNPQRRWAGTAGGPTPSAATGPGAKSLITRGQGGWPAAPSEGFAPKPTPT 150
 QY 46 CASRWPRGASRWPS 60
 DB 151 RNSSWPAASAPSGS 165

RESULT 9
 ABG03615
 ID ABG03615 standard; protein; 443 AA.
 XX
 AC ABG03615;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #3606.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.

XX WO200175067-A2.
 XX
 XX 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US008631.
 XX
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Dmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 XX N-PSDB; AAS67802.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity.
 XX
 XX Claim 20; SEQ ID NO 33974; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 XX sequences (II) is useful as hybridisation probes, polymerase chain
 XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 XX and in recombinant production of (II). The polynucleotides are also used
 XX in diagnostics as expressed sequence tags for identifying expressed
 XX genes. (I) is useful in gene therapy techniques to restore normal
 XX activity of (II) or to treat disease states involving (II). (II) is
 XX useful for generating antibodies against it, detecting or quantitating a
 XX polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 443 AA;

Query Match 16.6%; Score 75.5; DB 4; Length 443;
 Best Local Similarity 23.9%; Pred. No. 7.3;
 Matches 26; Conservative 7; Mismatches 41; Indels 35; Gaps 4;
 QY 5 PRRVSAGCGFADAHW-----TGLWTGLGEGGEGGIGPQGASPTPD 45
 DB 4 PRRQSECGAPTLTWPPGNGLPQQGASPLSASFGAGSGRGPAAAG-GSGASCTPSR 62
 QY 46 CASRWPRGASRWPS-----WSAGLT-----VRDRPQLGELCMGRG 79
 DB 63 GPASWSRGAQVPRSSRWAGSASXNAGSPTPTTSQPPRALCAAAG 111

RESULT 10
 ADC86843
 ID ADC86843 standard; protein; 1063 AA.
 XX
 AC ADC86843;
 XX
 DT 01-JAN-2004 (first entry)
 XX
 DE Human GPCR protein SEQ ID NO:1296.
 XX
 KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
 KW gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1270724-A2.
 XX
 PD 02-JAN-2003.
 XX
 XX 18-JUN-2002; 2002EP-00013517.
 XX
 XX 18-JUN-2001; 2001JP-00246789.
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.

XX Suwa M, Asai K, Akiyama Y, Aburatani H;
 XX
 XX WPI; 2003-315783/31.
 XX N-PSDB; ADC86842.
 XX
 XX New polynucleotide, useful for preparing a composition for treating a
 XX patient in need of increased or suppressed activity or expression of the
 XX guanosine triphosphate-binding protein coupled receptor.
 XX
 XX Claim 2; SEQ ID NO 1296; 28pp; English.
 XX
 XX The invention relates to a novel polynucleotide encoding a guanosine
 XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 XX the invention may have a use in gene therapy. The polynucleotide and
 XX polypeptide are useful for preparing a composition for treating a patient
 XX in need of increased or suppressed activity or expression of the
 XX guanosine triphosphate-binding protein coupled receptor. The protein
 XX sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.

SQ Sequence 1063 AA;

Query Match 16.6%; Score 75.5; DB 7; Length 1063;
 Best Local Similarity 28.8%; Pred. No. 19;
 Matches 32; Conservative 5; Mismatches 35; Indels 39; Gaps 8;

QY 2 CGRRP-----RVSAGCGFADAHWTGLTGL-GEQGGIGFEGQASPTP-----DC-- 46
 DB 164 CSVPREPCGGACRVIDGCG-SDA-----GPGMPTAASGVCGPHGRCVSPGGNFSIC 217

QY 47 -----ASRWPRASRW-----PWSAGLTVDRPQ-LGELQWGRG 79
 DB 218 DSGFTGTGYCHESEWPRTAGWGWAGLRPWLTPLASADIDDCLGQPCRNGG 268

RESULT 11

ADC86885
 ID ADC86885 standard; protein; 1356 AA.
 AC ADC86885;
 DT 01-JAN-2004 (first entry)
 DE Human GPCR protein SEQ ID NO:1338.
 XX human; GPCR; guanosine triphosphate-binding protein coupled receptor;
 XX Gene therapy.
 XX Homo sapiens.
 XX EPI270724-A2.
 XX 02-JAN-2003.
 XX 18-JUN-2002; 2002EP-00013517.
 XX 18-JUN-2001; 2001JP-00246799.
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;
 XX WPI; 2003-315783/31.
 XX N-PSDB; ADC86884.
 XX New polynucleotide, useful for preparing a composition for treating a
 PT patient in need of increased or suppressed activity or expression of the
 PT guanosine triphosphate-binding protein coupled receptor.
 XX Claim 2; SEQ ID NO 1338; 289p; English.

The invention relates to a novel polynucleotide encoding a guanosine
 triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 the invention may have a use in gene therapy. The polynucleotide and
 polypeptide are useful for preparing a composition for treating a patient
 in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.

SQ Sequence 1356 AA;

Query Match 16.6%; Score 75.5; DB 7; Length 1356;
 Best Local Similarity 28.8%; Pred. No. 24;
 Matches 32; Conservative 5; Mismatches 35; Indels 39; Gaps 8;

QY 2 CGRRP-----RVSAGCGFADAHWTGLTGL-GEQGGIGFEGQASPTP-----DC-- 46
 DB 164 CSVPREPCGGACRVIDGCG-SDA-----GPGMPTAASGVCGPHGRCVSPGGNFSIC 217

QY 47 -----ASRWPRASRW-----PWSAGLTVDRPQ-LGELQWGRG 79
 DB 218 DSGFTGTGYCHESEWPRTAGWGWAGLRPWLTPLASADIDDCLGQPCRNGG 268

RESULT 12

ABM65289
 ID ABM65289 standard; protein; 104 AA.
 AC ABM65289;
 DT 20-OCT-2003 (first entry)
 DE Propionibacterium acnes immunogenic polypeptide #29965.
 XX Acne vulgaris; antisborrheic; dermatological; antibacterial;
 XX immunostimulant; immune response; vaccine; immunogenic.
 XX Propionibacterium acnes.
 XX WO2003033515-A1.
 XX 24-APR-2003.
 XX 11-OCT-2002; 2002WO-US032727.
 XX 15-OCT-2001; 2001US-00978825.
 XX (CORI-) CORIXA CORP.
 XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 XX Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 XX Barth B, Vallieve-Douglass J;
 XX WPI; 2003-391789/36.
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.
 XX Claim 7; SEQ ID NO 29965; 1481pp; English.

The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 encoding a Propionibacterium acnes protein. The invention also relates to
 polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 immunogenic fragments of P. acnes polypeptides. The invention
 additionally encompasses expression vectors and host cells comprising a
 polynucleotide of the invention; antibodies against polypeptides of the
 CC polynucleotide of the invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridisation. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a specifically claimed P. acnes polypeptide which is
 CC thought to contain an immunogenic region. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 104 AA;

Query Match 16.3%; Score 74.5; DB 6; Length 104;
 Best Local Similarity 32.1%; Pred. No. 2;
 Matches 27; Conservative 7; Mismatches 35; Indels 15; Gaps 5;

QY 3 GPRPRVAGCGPADAHWTGLTGLGEGGQGGIGPBGQASPTPDCASRWPRASRWPSAG 62
 DB 14 GR-RRPACGCGSGHPRYEDPMPGLWRROE-AVAPEGHR-----CSSGFDSCTSGSVAAP 67
 QY 63 LTVDR-POLGE-----LCMG 77
 DB 68 FTVRSRVPTLSAPLRWALLCVG 91

RESULT 13
 ABP69309
 ID ABP69309 standard; protein; 156 AA.
 XX AC ABP69309;
 XX DT 20-JAN-2003 (first entry)
 XX DE Human polypeptide SEQ ID NO 1356.
 XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
 KW antiarthritic.
 XX OS Homo sapiens.
 XX PN WO200270539-A2.
 XX PD 12-SEP-2002.
 XX PF 05-MAR-2002; 2002WO-US005095.
 XX PR 05-MAR-2001; 2001US-00799451.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
 PI Wehrman T, Wang J, Wang D, Drmanac RT;
 XX DR WPI; 2002-759812/82.
 XX DR N-PSDB; ABZ11526.
 XX PT New polynucleotides comprising sequences assembled from expressed
 PT sequence tags (ESTs), useful for treating cell-proliferative,
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
 PT or coagulation disorders.
 XX PS Claim 9; SEQ ID NO 1356; 1012pp + Sequence Listing; English.
 XX CC The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence selected from any of 948 sequences (ABZ1119-
 CC ABZ12066) or their mature protein coding portion, active domain coding
 CC protein or complementary sequences. The polynucleotides are useful for
 CC identifying expressed genes or for physical mapping of human genome. The
 CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
 CC markers, as a food supplement, for generating antibodies, in medical
 CC imaging, screening and diagnostic assays and for treating cell-
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
 CC arthritis, etc. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 156 AA;

Query Match 16.3%; Score 74.5; DB 5; Length 156;
 Best Local Similarity 26.4%; Pred. No. 3;
 Matches 23; Conservative 7; Mismatches 20; Indels 37; Gaps 3;
 QY 10 AGCGFADAHWTG-----LWTGL-----GEGQGG 33
 DB 13 AGCGFLGVYVGVASCLREHAFVLVANATHYGSAGALTATALTGVCLGERGROAG 72
 QY 34 IGPEGQASPTPDCASRWPRASRW-PW 59
 DB 73 GWREGFAGEGDLRRRGPAASRWVPM 99

RESULT 14
 ABG14843
 ID ABG14843 standard; protein; 180 AA.
 XX AC ABG14843;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #14834.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US008631.
 XX PR 31-MAR-2000; 2000US-00540217.
 XX PR 23-AUG-2000; 2000US-00649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS79030.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX PS Claim 20; SEQ ID NO 45202; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC patent did not appear in the printed specification. Note: The sequence data for this
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX


```

SQ Sequence 180 AA;
Query Match 16.3%; Score 74.5; DB 4; Length 180;
Best Local Similarity 26.3%; Pred. No. 3.5;
Matches 31; Conservative 6; Mismatches 40; Indels 41; Gaps 6;
QY 2 CGRRPRVS-AGC-----GFADAHTGLWTG-----LGEQGGGIG 35
DB 44 CWRTRASCAPCAGATAPRAARTRSSGWCSPWSSWPASCPTTCCWCAGTGRPAATS 103
QY 36 PEGQASPT--PDC-----ASRWPRAS-----RWPWAGLTVRDRPOLGELCMGR 78
DB 104 PRAFSTPTTSPSCSPASTASPTPCSTASSARPPTGTWPSAGPWPSPAPGPGFGR 161

```

RESULT 15

ABG12827
ID ABG12827 standard; protein; 180 AA.

XX AC ABG12827;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #12818.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS77014.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX PS Claim 20; SEQ ID NO 43186; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: the sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 180 AA;

Query Match 16.3%; Score 74.5; DB 4; Length 180;
Best Local Similarity 26.3%; Pred. No. 3.5;
Matches 31; Conservative 6; Mismatches 40; Indels 41; Gaps 6;

```

QY 2 CGRRPRVS-AGC-----GFADAHTGLWTG-----LGEQGGGIG 35
DB 44 CWRTRASCAPCAGATAPRAARTRSSGWCSPWSSWPASCPTTCCWCAGTGRPAATS 103
QY 36 PEGQASPT--PDC-----ASRWPRAS-----RWPWAGLTVRDRPOLGELCMGR 78
DB 104 PRAFSTPTTSPSCSPASTASPTPCSTASSARPPTGTWPSAGPWPSPAPGPGFGR 161

```

Search completed: August 13, 2004, 09:06:09

Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2004, 09:05:07 ; Search time 19 Seconds
(without alignments)
214.655 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456

Sequence: 1 MCGRRPRVSGCGFADAHWT.....SAGLTVDRPQLGELCMGRG 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgm2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgm2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgm2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgm2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgm2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgm2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86.5	19.0	326	4	US-09-252-991A-21802
2	78.5	17.2	451	4	US-09-252-991A-27602
3	75.5	16.6	423	4	US-09-252-991A-24347
4	75.5	16.6	663	4	US-09-252-991A-30843
5	75	16.4	431	4	US-09-252-991A-26751
6	74.5	16.3	312	4	US-09-252-991A-31204
7	73.5	16.1	170	4	US-09-199-637A-299
8	73	16.0	138	4	US-09-252-991A-24447
9	73	16.0	436	4	US-09-252-991A-33064
10	72.5	15.9	148	4	US-09-252-991A-17048
11	72.5	15.9	719	4	US-09-252-991A-30052
12	71	15.6	330	4	US-09-252-991A-25986
13	71	15.6	860	4	US-09-252-991A-25911
14	70.5	15.5	141	4	US-09-252-991A-16951
15	70	15.4	264	4	US-09-252-991A-24262
16	70	15.4	532	1	US-08-494-168-9
17	69	15.1	153	4	US-09-252-991A-21201
18	69	15.1	178	4	US-09-252-991A-23713
19	68	15.1	1217	4	US-09-252-991A-26104
20	68	14.9	149	4	US-09-252-991A-25693
21	68	14.9	169	4	US-09-252-991A-22543
22	67.5	14.8	121	4	US-09-072-596-353
23	67.5	14.8	121	4	US-09-072-967-258
24	67.5	14.8	203	4	US-09-252-991A-21364
25	67.5	14.8	239	4	US-09-252-991A-28985
26	67.5	14.8	676	4	US-09-252-991A-23181
27	67.5	14.8	1958	1	US-07-945-283-2

28	67	14.7	223	4	US-09-252-991A-20332	Sequence 20332, A
29	67	14.7	372	4	US-09-252-991A-20108	Sequence 20108, A
30	67	14.7	475	4	US-09-252-991A-22247	Sequence 22247, A
31	67	14.7	728	4	US-09-252-991A-23613	Sequence 23613, A
32	66.5	14.6	117	4	US-09-252-991A-20556	Sequence 20556, A
33	66.5	14.6	196	4	US-09-252-991A-19503	Sequence 19503, A
34	66.5	14.6	401	4	US-09-252-991A-21672	Sequence 21672, A
35	66	14.5	192	4	US-09-252-991A-27287	Sequence 27287, A
36	66	14.5	206	4	US-09-252-991A-31481	Sequence 31481, A
37	66	14.5	274	4	US-09-252-991A-23172	Sequence 23172, A
38	66	14.5	371	4	US-09-252-991A-29793	Sequence 29793, A
39	66	14.5	518	4	US-09-252-991A-23604	Sequence 23604, A
40	66	14.5	533	4	US-09-252-991A-23560	Sequence 23560, A
41	65.5	14.4	112	4	US-09-489-039A-7360	Sequence 7360, Ap
42	65.5	14.4	125	4	US-09-252-991A-27936	Sequence 27936, A
43	65.5	14.4	133	4	US-09-252-991A-27120	Sequence 27120, A
44	65.5	14.4	193	4	US-09-252-991A-28920	Sequence 28920, A
45	65.5	14.4	389	2	US-08-485-449-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-252-991A-21802
; Sequence 21802, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21802
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21802

Query Match 19.0%; Score 86.5; DB 4; Length 326;
Best Local Similarity 42.1%; Pred. No. 0.055;
Matches 24; Conservative 3; Mismatches 27; Indels 3; Gaps 2;
Qy 4 RPRRVSG--CGFADAHWTGLTGLGEGGEGGIGPEGQASPTPDCASRWPRASRWMP 58
Db 261 RNRPRAGAACAPAI PHSITRIATGSSRQWAGKAAPV-RAGRRPGGASRWPRPQRRWP 316

RESULT 2

US-09-252-991A-27602
; Sequence 27602, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27602
; LENGTH: 451
; TYPE: PRT

```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27602

Query Match      17.2%; Score 78.5; DB 4; Length 451;
Best Local Similarity 34.1%; Pred. No. 0.61;
Matches 28; Conservative 3; Mismatches 38; Indels 13; Gaps 3;

QY 5 PRRVSAGCGFADAHWTGLGEGGEGIGPEGQASPTTDCASRWPRSRWPSAGLT 64
DB 142 PRLAFGAAGROPAAH-----GAGEKQFADRLPAAQRTPRPACAGPATCAERDPGPGRP 195

QY 65 VR-----DRPO--LGELCMGRG 79
DB 196 VRRPAGAERPORPAGQAPGGRG 217

RESULT 3
US-09-252-991A-24347
; Sequence 24347, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24347
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24347

Query Match      16.6%; Score 75.5; DB 4; Length 423;
Best Local Similarity 35.2%; Pred. No. 1.2;
Matches 19; Conservative 6; Mismatches 14; Indels 15; Gaps 2;

QY 40 ASPTTDCASRWPRSRASRW-----WSAGLTVRDPQGLGELCMGRG 79
DB 3 ASPSSVWASRW-RSPTRWPTSSRTTTTPAAFSNWTACACRPRPRKEAPLSRG 55

RESULT 4
US-09-252-991A-30843
; Sequence 30843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30843
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30843

Query Match      16.6%; Score 75.5; DB 4; Length 663;
Best Local Similarity 55.6%; Pred. No. 2;
Matches 15; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

QY 35 GPEGQASPTTDCASRWPRSRASRWPSA 61
DB 110 GP-SPASPAPAAACRRWPRAACHWPASA 135

RESULT 5
US-09-252-991A-26751
; Sequence 26751, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26751
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26751

Query Match      16.4%; Score 75; DB 4; Length 431;
Best Local Similarity 30.0%; Pred. No. 1.4;
Matches 27; Conservative 6; Mismatches 27; Indels 30; Gaps 5;

QY 2 CGRRVVSAGCGFADA-----HWTGLWTGLGEGGEGIGPEGQASPTTDC-ASRWP 51
DB 12 CDRPCTAGTPTSWRTAPDRAPARWARGSWPGRRPGRHRRWPPE-----PGCRSRRP 65

QY 52 RSAS-----RWP-----WSAGLTVRD 67
DB 66 RSALAPGEADRAPWRWPDACRWTAPSLRD 95

RESULT 6
US-09-252-991A-31204
; Sequence 31204, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31204
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31204

Query Match      16.3%; Score 74.5; DB 4; Length 312;
Best Local Similarity 37.9%; Pred. No. 1.1;
Matches 22; Conservative 2; Mismatches 21; Indels 13; Gaps 2;

QY 4 RRRVVSAGCGFADAHWTGLGEGGEGIGPEGQASPTTDCASRWPRSRASRWPSA 61
DB 210 RPRASHRGIS-----GRGAGSGPADVCPAPDAAGR--RGAIRSPGGA 254

RESULT 7
US-09-199-637A-299
```

```
; Sequence 299, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 299
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-299

Query Match 16.1%; Score 73.5; DB 4; Length 170;
Best Local Similarity 32.1%; Pred. No. 0.75;
Matches 18; Conservative 5; Mismatches 20; Indels 13; Gaps 1;

Y 3 GRPRVSAGCGFADAHWTGLTGLGEGGEGIGPEGQASPTPDCASRWPSASRWP 58
D 16 GWPRMPDAGWRRACWRRRT-----AWACPATNCWGRRRSARRP 58

RESULT 8
US-09-252-991A-24447
; Sequence 24447, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24447
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24447

Query Match 16.0%; Score 73; DB 4; Length 138;
Best Local Similarity 32.4%; Pred. No. 0.68;
Matches 23; Conservative 8; Mismatches 30; Indels 10; Gaps 4;

Y 6 RVVSAGCGFADAHWTG--LWTGLGEGQ-----EGGIGPEGOAS--PTPDCAS--RWPSAS 55
D 31 RSAAPGSPASAPAWPGAAVWRRSGTGTQRSGRSRPSGRAGRSAPRGRSRPQSV 90
Y 56 RWPWSAGLTVR 66
D 91 GWPGPARYSSR 101

RESULT 9
US-09-252-991A-33064
; Sequence 33064, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 33064
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33064

Query Match 16.0%; Score 73; DB 4; Length 436;
Best Local Similarity 33.3%; Pred. No. 2.4;
Matches 24; Conservative 6; Mismatches 20; Indels 22; Gaps 3;

Y 6 RVVSAGCGFADAHWTGLTGLGEGGEGIGPEGQASPTPDCASRWPSASRWPWSA--- 61
D 255 RVVAA-----QLPVGLEAGQGQGGPQLQRAEPGTDAASQ-----PWQAPGV 296
Y 62 GLTVDRDPOLGE 73
D 297 GVLEQEAUVGE 308

RESULT 10
US-09-252-991A-17048
; Sequence 17048, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17048
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17048

Query Match 15.9%; Score 72.5; DB 4; Length 148;
Best Local Similarity 31.2%; Pred. No. 0.83;
Matches 20; Conservative 3; Mismatches 30; Indels 11; Gaps 3;

Y 2 CGPRRVSAGC-----GFADAHWTGLTGLGEGGEGIGPEGOASPTPDCASRWP--- 53
D 62 CGTPRATPCNCTAAAGNTSASWSRSARMSASNSASSP---ASPTPCVSTWGXACRA 118
Y 54 ASRW 57
D 119 ASGW 122

RESULT 11
US-09-252-991A-30052
; Sequence 30052, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 30052

; LENGTH: 719

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30052

Query Match 15.9%; Score 72.5; DB 4; Length 719;

Best Local Similarity 26.7%; Pred. No. 4.8;

Matches 28; Conservative 10; Mismatches 38; Indels 29; Gaps 5;

QY 3 GRPRVSAGCGFADAHWTGLTGLGEGEGG-----GPEGASPT-----PDCASR 49

DB 11 GPRRRRGCGACRACVPVA-WSGVGSAPRGVLRPPAALRRGGDRQAATGRRAPGCAAR 69

QY 50 WPRSASR-----WPW-----SAGLTVDRPQGLGELCMGRG 79

DB 70 WRPAAGRAIAATGAAVPGVAPRGAAAGVPAASDRRRARAPGAG 114

RESULT 12

US-09-252-991A-25986

; Sequence 25986, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 25986

; LENGTH: 330

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-25986

Query Match 15.6%; Score 71; DB 4; Length 330;

Best Local Similarity 29.7%; Pred. No. 3;

Matches 30; Conservative 3; Mismatches 22; Indels 46; Gaps 5;

QY 4 RPRVSAGCGFADAHWTGLTGLGEGEG-----GPEGASPT 43

DB 228 RPRPGEG-----OPGLRGQEGPRGRQGRQGRPDRAPEGARRAGQRRPRFP 276

QY 44 PDCASRWPRASRPW-----SAGLTVDRP 69

DB 277 PDRFAGDPRHRTGPWADAPDRGQRAESAGLEADLRDQP 317

RESULT 13

US-09-252-991A-25911

; Sequence 25911, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 25911

; LENGTH: 860

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-25911

Query Match 15.6%; Score 71; DB 4; Length 860;

Best Local Similarity 40.5%; Pred. No. 8.6;

Matches 17; Conservative 3; Mismatches 18; Indels 4; Gaps 1;

QY 36 PEGQASPTPDCASRWPRASRWPMWSAGLTVDRPQGLGELCMG 77

DB 781 PPSSVSPTPASMTWRRTATRWSSAD-----SRPQITILAE 818

RESULT 14

US-09-252-991A-16951

; Sequence 16951, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 16951

; LENGTH: 141

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16951

Query Match

Best Local Similarity 15.5%; Score 70.5; DB 4; Length 141;

Matches 19; Conservative 5; Mismatches 18; Indels 25; Gaps 3;

QY 36 PEGQASPTPDCAS-----RWP-----RSASRWPMWSAGLTVDRPQ--- 70

DB 27 PSRGAGAPARGASWMPGPGFQWPGPSRSGAPRRHAGRGARWPCGCTARGWPRSPC 86

QY 71 LGELCMG 77

DB 87 LPRRCPG 93

RESULT 15

US-09-252-991A-24262

; Sequence 24262, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

```
; SEQ ID NO 24262
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24262

Query Match      15.4% Score 70; DB 4; Length 264;
Best Local Similarity 35.5%; Pred. No. 3;
Matches 22; Conservative 5; Mismatches 25; Indels 10; Gaps 3;

QY      5 PRRYSAGCGFADAHWTGLWTGLGEGGIGPEGQASPTPDCASRWPRSASRWPSAGLT 64
Db      100 PLRAGAVAGSA-----ARLGHADQPLRG-KGRDRWFCCARGWP-TASCWTWSAGRS 149

QY      65 VR 66
Db      150 SR 151
```

Search completed: August 13, 2004, 09:08:19
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2004, 09:06:12 ; Search time 46 Seconds
(without alignments)
539.136 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456

Sequence: 1 MCGRRPRVSAGCGFADAHWT.....SAGLTVDRLQGLCLMGRG 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	79	12	US-09-972-032-2
2	79.5	17.4	123	16	US-10-437-963-125319
3	79	17.3	99	16	US-10-437-963-175071
4	79	17.3	813	16	US-10-437-963-175071
5	78	17.1	11088	16	US-10-203-295-7
6	77	16.9	215	15	US-10-094-749-1759
7	77	16.9	235	12	US-10-343-650A-98
8	75.5	16.6	146	16	US-10-437-963-190988
9	75.5	16.6	215	12	US-10-425-114-70579
10	75.5	16.6	231	16	US-10-437-963-134598
11	75.5	16.6	1063	14	US-10-017-161-1624
12	75.5	16.6	1063	15	US-10-292-798-1296
13	75.5	16.6	1356	14	US-10-017-161-1678
14	75.5	16.6	1356	15	US-10-292-798-1338
15	73.5	16.1	170	10	US-09-975-719-299

16	73	16.0	515	16	US-10-437-963-187892	Sequence 187892,
17	72	15.8	514	12	US-10-296-115-1432	Sequence 1432, Ap
18	71	15.6	253	11	US-09-833-245-1132	Sequence 1132, Ap
19	71	15.6	267	14	US-10-156-761-10080	Sequence 10080, A
20	71	15.6	625	14	US-10-156-761-15008	Sequence 15008, A
21	71	15.6	1081	16	US-10-437-963-196915	Sequence 196915,
22	70	15.4	177	12	US-10-425-114-42696	Sequence 42696, A
23	70	15.4	182	12	US-10-425-114-51184	Sequence 51184, A
24	70	15.4	222	12	US-10-425-114-64974	Sequence 64974, A
25	70	15.4	569	14	US-10-156-761-14513	Sequence 14513, A
26	70	15.4	677	14	US-10-259-165-192	Sequence 192, App
27	70	15.4	1214	12	US-10-267-502-333	Sequence 333, App
28	70	15.4	1669	12	US-09-918-715-252	Sequence 252, App
29	70	15.4	1869	15	US-10-372-683-8	Sequence 8, Appli
30	69.5	15.2	552	14	US-10-106-698-6365	Sequence 6365, Ap
31	69	15.1	122	16	US-10-437-963-127558	Sequence 127558,
32	69	15.1	141	16	US-10-437-963-127488	Sequence 127488,
33	69	15.1	190	12	US-10-425-114-72793	Sequence 72793, A
34	68.5	15.0	125	16	US-10-437-963-186225	Sequence 186225,
35	68.5	15.0	179	16	US-10-437-963-176622	Sequence 176622,
36	68.5	15.0	207	15	US-10-104-047-2751	Sequence 2751, Ap
37	68.5	15.0	207	16	US-10-471-115-7	Sequence 7, Appli
38	68.5	15.0	236	16	US-10-437-963-178641	Sequence 178641,
39	68.5	15.0	2169	9	US-09-738-626-5455	Sequence 5455, Ap
40	68	14.9	176	15	US-10-108-260A-2704	Sequence 2704, Ap
41	68	14.9	280	12	US-10-425-114-45512	Sequence 45512, A
42	68	14.9	304	12	US-10-399-645-5	Sequence 5, Appli
43	68	14.9	316	12	US-10-262-511-118	Sequence 118, App
44	67.5	14.8	121	14	US-10-193-002-253	Sequence 253, App
45	67.5	14.8	121	14	US-10-084-843-258	Sequence 258, App

ALIGNMENTS

RESULT 1

US-09-972-032-2
; Sequence 2, Application US/09972032
; Publication No. US20020086361A1
; GENERAL INFORMATION:
; APPLICANT: Case Western Reserve University
; APPLICANT: Montano, Monica
; APPLICANT: Sutton, Amelia
; TITLE OF INVENTION: A Modulator of Antiestrogen Pharmacology
; FILE REFERENCE: 27708/04003
; CURRENT APPLICATION NUMBER: US/09/972,032
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/238,190
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-032-2

Query Match	100.0%	Score	456	DB	12	Length	79
Best Local Similarity	100.0%	Pred. No.	1.7e-37				
Matches	79	Conservative	0	Mismatches	0	Indels	0
Qy	1	MCGRRPRVSAGCGFADAHWTGLTGLGEGQBGIGEGQASPTPDCA	SWPSASRPWS	60			
Db	1	MCGRRPRVSAGCGFADAHWTGLTGLGEGQBGIGEGQASPTPDCA	SWPSASRPWS	60			
Qy	61	AGLTVDRLQGLCLMGRG	79				
Db	61	AGLTVDRLQGLCLMGRG	79				

RESULT 2

US-10-437-963-125319
; Sequence 125319, Application US/10437963


```
; SEQ ID NO 7
; LENGTH: 11088
; TYPE: PRT
; ORGANISM: Streptomyces noursei ATCC 11455
US-10-203-295-7

Query Match      17.1%; Score 78; DB 16; Length 11088;
Best Local Similarity 37.2%; Pred. No. 2.9e+02;
Matches 32; Conservative 7; Mismatches 33; Indels 14; Gaps 6;

Qy 4 RPRVSGCGFADAHWTGL---WTGLGEGGEGGIGPEGQASPT-PDCASR-WPRS--ASR 56
Db 828 RPELSAVTGARAHVRGVTWRAGLFD---GTGARRADLPTTFQHQRPWPTAARAAQ 883

Qy 57 WPSAGLTVDRPQLG---ELCMGRG 79
Db 884 DVTAAGLGAADHPLLGATVELADGAG 909

RESULT 6
US-10-094-749-1759
; Sequence 1759, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUKIO
; APPLICANT: HIG, YURI
; APPLICANT: OTSUKA, KAGURU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYU
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1759
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1759

Query Match      16.9%; Score 77; DB 15; Length 215;
Best Local Similarity 31.9%; Pred. No. 8.4;
Matches 23; Conservative 3; Mismatches 28; Indels 18; Gaps 4;

Qy 7 RVSGCGFADAHWTGL-----WTGLGEGGEGGIGPEGQASPTPD-CASRWPRSASRW 57
Db 64 RCGARC-----WEGSSLAQLQPPWTPSGSLTAAGVQVAD---SCWRSWRSRSCFRW 114

Qy 58 PWSAGLTVDRRP 69
Db 115 PPSSTAARSSP 126

RESULT 7
US-10-343-650A-98

; Sequence 98, Application US/10343650A
; Publication No. US20040067499A1
; GENERAL INFORMATION:
; APPLICANT: HAGA, TATSUYA
; TITLE OF INVENTION: NOVEL G-PROTEIN COUPLED RECEPTOR
; FILE REFERENCE: 31671-186347
; CURRENT APPLICATION NUMBER: US/10/343,650A
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: JP 2000/237818
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: JP 2001/34434
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 694
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 98
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-343-650A-98

Query Match      16.9%; Score 77; DB 12; Length 235;
Best Local Similarity 37.2%; Pred. No. 9.2;
Matches 16; Conservative 3; Mismatches 18; Indels 6; Gaps 1;

Qy 19 WTGLWTGLGEGGIGPEGQASPTPD-CASRWPRSASRWPSA 61
Db 11 WLGLWVGL-----GLRPTFRVCSPLCGPLWPRASLVCWGS 47

RESULT 8
US-10-437-963-190988
; Sequence 190988, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 190988
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(146)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_8734C.1.pep
US-10-437-963-190988

Query Match      16.6%; Score 75.5; DB 16; Length 146;
Best Local Similarity 31.9%; Pred. No. 8.2;
Matches 22; Conservative 5; Mismatches 19; Indels 23; Gaps 3;

Qy 3 GRPRRYVSGCGFADAHWTGLWTGLGEGGEGGIGPEGQASPTP----DCASRWPR----- 52
Db 9 GRPRRAGXXYG-----GRGPRGVLGPGQAPRPHGRASPVARRRRLPLV 57

Qy 53 --SASRWPW 59
Db 58 PLSLESSPW 66
```


QY 47 -----ASRWPGRASRW-----PWSAGLTVDRPQ-LGELCMGRG 79
DB 218 DSGFTGYCHSEWPRTAGWWGAGLRPWLTPPLASADIDDCLGQPCRNGG 268

RESULT 13

US-10-017-161-1678
; Sequence 1678, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: ABURATANI, HIROYUKI
; APPLICANT: AKIYAMA, YUTAKA
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1678
; LENGTH: 1356
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1678

Query Match 16.6%; Score 75.5; DB 14; Length 1356;
Best Local Similarity 28.8%; Pred. No. 68;
Matches 32; Conservative 5; Mismatches 35; Indels 39; Gaps 8;
QY 2 CGRPR-----RVSGCCGADAHWTGLWTGL-GEQGGIGPEGQASPTP-----DC-- 46
DB 164 CSVPREPCGGACRVIDGCG-SDA-----GPGMPTAASGVCGPHGRCVSPGQGNFSCIC 217
QY 47 -----ASRWPGRASRW-----PWSAGLTVDRPQ-LGELCMGRG 79
DB 218 DSGFTGYCHSEWPRTAGWWGAGLRPWLTPPLASADIDDCLGQPCRNGG 268

RESULT 14

US-10-292-798-1338
; Sequence 1338, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1338
; LENGTH: 1356
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1338

Query Match 16.6%; Score 75.5; DB 15; Length 1356;
Best Local Similarity 28.8%; Pred. No. 68;
Matches 32; Conservative 5; Mismatches 35; Indels 39; Gaps 8;
QY 2 CGRPR-----RVSGCCGADAHWTGLWTGL-GEQGGIGPEGQASPTP-----DC-- 46

DB 164 CSVPREPCGGACRVIDGCG-SDA-----GPGMPTAASGVCGPHGRCVSPGQGNFSCIC 217
QY 47 -----ASRWPGRASRW-----PWSAGLTVDRPQ-LGELCMGRG 79
DB 218 DSGFTGYCHSEWPRTAGWWGAGLRPWLTPPLASADIDDCLGQPCRNGG 268

RESULT 15

US-09-975-719-299
; Sequence 299, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 299
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-299

Query Match 16.1%; Score 73.5; DB 10; Length 170;
Best Local Similarity 32.1%; Pred. No. 15;
Matches 18; Conservative 5; Mismatches 20; Indels 13; Gaps 1;
QY 3 GRPRVSGCGFADAHWTGLWTGLGEGGIGPEGQASPTPDCASRWPGRSASRWP 58
DB 16 GWPRVMPAGWRRACWRRWT-----AWACPATNCWSGWRRSARP 58

Search completed: August 13, 2004, 09:09:17
Job time : 47 secs

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
A;Genetics:
A;Gene: XF0267
C;Superfamily: Autotransporter subtilisin-like protease precursor; subtilisin homology

Query Match 15.5%; Score 70.5; DB 2; Length 905;
Best Local Similarity 27.2%; Pred. No. 15;
Matches 22; Conservative 8; Mismatches 22; Indels 29; Gaps 5;
QY 7 RVSAACGCGF-----ADAHWTGL-WTGLGEGQ-----EGIGPEGQASP 42
Db 786 REQSGSGFLQVQSAQATRSQLLAGVRTWEGWAGVQWRYGEWQTLRQSLNPAQSFTA 845
QY 43 TPDCASRW-PRASRWPNWSAG 62
Db 846 T-----SSWTPLVASSWPGRSG 862

RESULT 3
H8286
serine proteinase XP0267 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 24-Nov-2003
C;Accession: H8286
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H8286
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-910 <SIM>
A;Cross-references: GB:A8003880; GB:A8003849; NID:9105080; PIDN:APF83080.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
A;Genetics:
A;Gene: XF0267
C;Superfamily: Autotransporter subtilisin-like protease precursor; subtilisin homology

Query Match 15.5%; Score 70.5; DB 2; Length 910;
Best Local Similarity 27.2%; Pred. No. 15;
Matches 22; Conservative 8; Mismatches 22; Indels 29; Gaps 5;
QY 7 RVSAACGCGF-----ADAHWTGL-WTGLGEGQ-----EGIGPEGQASP 42
Db 791 REQSGSGFLQVQSAQATRSQLLAGVRTWEGWAGVQWRYGEWQTLRQSLNPAQSFTA 850
QY 43 TPDCASRW-PRASRWPNWSAG 62
Db 851 T-----SSWTPLVASSWPGRSG 867

RESULT 4
CGHU4B
collagen alpha 1(IV) chain precursor - human
N;Alternate names: procollagen alpha 1(IV) chain
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 07-Dec-1999
C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S35614; A02863; A59
J. Biol. Chem. 264, 13565-13571, 1989
R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
A;Title: Structural organization of the gene for the alpha-1 chain of human type IV colla
A;Reference number: S16876; MUID:89340433; PMID:2701944
A;Accession: S16876
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1669 <SO11>
A;Cross-references: EMBL:J04217; GB:J05039; NID:gl80800; PIDN:AA53098.1; PID:gl80803
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
R;Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A;Reference number: A92690; MUID:89034231; PMID:3182844
A;Accession: A32117
A;Molecule type: DNA
A;Residues: 1-28 <SO12>
A;Cross-references: EMBL:J04217; NID:gl80759; PIDN:AAA53097.1; PID:G553233
R;Roesschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane co
A;Reference number: S02738; MUID:89030632; PMID:2846280
A;Accession: S02738
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-6, 'L', 8-28 <POB>
A;Cross-references: EMBL:X12784; NID:G30072
R;Brace, D.; Oberbauer, I.; Bieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;
Eur. J. Biochem. 168, 529-536, 1987
A;Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem
A;Reference number: S00048; MUID:88029471; PMID:3311751
A;Accession: S00048
A;Molecule type: mRNA
A;Residues: 1-318, 'A', 320-944 <BRAL>
A;Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067
A;Accession: S25826
A;Molecule type: Protein
A;Residues: 271-318, 'A', 320-554 <BRA2>
R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S c
A;Reference number: A23115; MUID:86004708; PMID:4043082
A;Accession: A23115
A;Molecule type: Protein
A;Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>
A;Experimental source: placenta
A;Note: the amino end of the mature form is blocked
R;Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.
FEBS Lett. 225, 188-194, 1987
A;Title: Complete primary structure of the alpha(1)-chain of human basement membrane (7S
A;Reference number: S00207; MUID:88083584; PMID:3691802
A;Accession: S00207
A;Molecule type: mRNA
A;Residues: 244-530 <SO13>
A;Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549
R;Eble, J.A.; Golbak, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen n
A;Reference number: S39614; MUID:94038963; PMID:8223488
A;Accession: S39614
A;Molecule type: Protein
A;Residues: 371-554 <EBL>
R;Babel, W.; Glanville, R.W.
Eur. J. Biochem. 143, 545-556, 1984
A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid sec
A;Reference number: A02863; MUID:85003629; PMID:6434307

A:Accession: A02863
A:Molecule type: Protein
A:Residues: 534-718, D⁷²⁰⁻⁸³⁶, Y⁸³⁸⁻⁸⁴¹, P⁸⁴³⁻⁹⁰³, Q⁹⁰⁵⁻⁹¹³, K⁹¹⁵⁻⁹⁹⁷, K⁹⁹⁹
A:Experimental source: Placenta
R:Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A:Title: Pepsin fragments of human placental basement-membrane collagens showing interrupted
A:Reference number: S16908; MUID:82005835; PMID:6792033
A:Accession: A58517
A:Molecule type: Protein
A:Residues: 534-537, G⁵³⁹, G⁵⁴¹⁻⁵⁴², X⁵⁴⁴⁻⁵⁵³, 1389-1405, XX¹⁴⁰⁸⁻¹⁴⁰⁹, X¹⁴¹¹⁻¹⁴¹⁴
R:Macwright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
Biochemistry 22, 4940-4948, 1983
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (b
A:Reference number: S16910; MUID:84053346; PMID:6416291
A:Accession: S16910
A:Molecule type: Protein
A:Residues: 534-537, G⁵³⁹, G⁵⁴¹⁻⁵⁴², G⁵⁴⁴⁻⁵⁴⁹, 939-940, M⁹⁴²⁻⁹⁴⁴, V⁹⁴⁶, X⁹⁴⁸
A:Experimental source: Placenta
R:Pihlajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; F
J. Biol. Chem. 260, 7681-7687, 1985
A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
A:Reference number: S01466; MUID:85207819; PMID:2581969
A:Accession: S01466
A:Molecule type: mRNA
A:Residues: 1256-1669 <PIH>
A:Cross-references: EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PID:G180424
R:Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A:Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
A:Reference number: S16879; MUID:85216555; PMID:2582422
A:Accession: S16879
A:Molecule type: mRNA
A:Residues: 1259-1669 <BRI>
A:Cross-references: EMBL:M11315; NID:G180817; PIDN:AAA52042.1; PID:G180818
R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
Eur. J. Biochem. 147, 217-224, 1985
A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A:Reference number: A02864; MUID:85127033; PMID:2578961
A:Accession: S19091
A:Molecule type: Protein
A:Residues: 1435-1461, H¹⁴⁶³⁻¹⁴⁸², X¹⁴⁸⁴⁻¹⁴⁹¹, 1501-1514, X¹⁵¹⁶⁻¹⁵¹⁹, 1534-1553, X¹⁵⁵⁵
R:Stebold, B.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
A:Reference number: S02550; MUID:89005112; PMID:2844531
A:Contents: annotation; disulfide bonds
C:Genetics:
A:Gene: GDB:COL4A1
A:Cross-references: GDB:119791; OMIM:120130
A:Map position: 13q34-13q34
A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/3
A:Exons: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/3
A:Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(IV)
A:Associations: among trimer amino-terminal domains (disulfide and desmosine cross-links), dim
A:Function: associations in the interrupted helical domain (with disulfide and desmosine cr
A:Function: structural component of extracellular basement membrane
A:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F:29-162/Domain: amino-terminal nonhelical, 7S <7SD>
F:163-1440/Domain: interrupted helical <COL>
F:414-452/Region: integrin binding #status experimental
F:597-599/Region: cell attachment (R-G-D) motif
F:917-919/Region: cell attachment (R-G-D) motif
F:968-970/Region: cell attachment (R-G-D) motif
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:27/Modified site: blocked amino end (A1a) (in mature form) #status experimental
F:31,36,39,41,125,434,467,470/disulfide bonds: interchain #status predicted

F:45,48,78,90,129,156,172,217,228,231,277,295,298,322,343,361,460,463,497,527,540,543,573,
1081,1084,1089,1117,1132,1150,1165,1182,1185,1188,1206,1235,1255,1283,1304,1319,1328,134,
F:45,48,78,90,129,156,217,228,231,277,295,298,322,343,361,460,463,497,527,543,573,582,61,
99,1117,1132,1150,1165,1182,1185,1188,1206,1235,1255,1283,1304,1319,1328,1340,1356,1371,
F:54,63,75,84,87,96,102,105,108,111,117,120,123,138,141,147,150,153,159,167,178,181,184,
F:419,422,425,439,445,448,451,479,485,491,494,503,512,518,524,530,546,549,552,555,561,567,
F:745,748,751,754,763/Modified site: 4-hydroxyproline (Pro) #status experimental
F:126/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:129/Modified site: allysine (Lys) #status predicted
F:172,540,947/Modified site: 5-hydroxylysine (Lys) #status atypical
F:272,645/839/Modified site: 4-hydroxyproline (Pro) #status atypical
F:446-447/Cleavage site: Gly-Ile (gelatinase B) #status predicted
F:766,775,784,787,790,796,799,804,810,816,822,834,860,863,869,872,875,887,890,893,899,902,
23,1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,1256,
F:1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,1256,
F:1437/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1120,1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F:1120,1268/Binding site: carbohydrate (Lys) (covalent) (partial) #status experimental
F:1214,1424/Modified site: 3-hydroxyproline (Pro) #status absent
F:1392,1395,1398,1404/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1460-1548,1493-1551/Disulfide bonds: (Or 1460-1551, 1493-1548) #status predicted
F:1505-1511,1616-1622/Disulfide bonds: #status predicted
F:1570-1662,1604-1665/Disulfide bonds: (Or 1570-1665, 1604-1662) #status predicted

Query Match 15.4%; Score 70; DB 1; Length 1669;
Best Local Similarity 31.1%; Pred. No. 31;
Matches 23; Conservative 7; Mismatches 24; Indels 20; Gaps 3;
QY 10 AGCGFADAHWT-----GLWTGLG---EGQGGGGGPGQASPTPD-----CAGR 49
Db 34 SCGCKCCHGVGKQKGERGLQGVGPGMGQGGPGQKQKGTGTR 93
QY 50 WPRSASRWPMWSAGL 63
Db 94 GPGGASGYPGNPGL 107

RESULT 5
T35294
Probable endo alpha-1,4 polygalactosaminidase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21574
A:Accession: T35294
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-282 <SEE>
A:Cross-references: EMBL:AL096872; PIDN:CAB51262.1; GSPDB:GN00070; SCOEDB:SC5F7.23C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5F7.23C

Query Match 14.9%; Score 68; DB 2; Length 282;
Best Local Similarity 37.3%; Pred. No. 9;
Matches 19; Conservative 3; Mismatches 25; Indels 4; Gaps 2;
QY 10 AGCGFADAHWTGLWTGLGQGGGGGPGQASPTPDCA SRW-PRASRWPMW 59
Db 16 AGCTAPGDDGGVGGVGGGGGGGGGGVGT---AGHWRTPTGTATQW 63

RESULT 6
B45344
Probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)
C:Species: suid herpesvirus 1
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
R:Vlcek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwytzer, M.
Virology 179, 385-377, 1990
A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented op
A:Reference number: A45344; MUID:91021039; PMID:2171211

```

Best Local Similarity   29.8%; Pred. No.17;
Matches    14; Conservative      6; Mismatches    25; Indels     2; Gaps     1;

QY       31 EGGIGPEQAASPTPCASRWPP--RSASRWPWSAGLTVDRPQLGELC 75
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB        343 KGGPGRCGCPLPDATHNFFVRQLVTNTGWGTGLDIRPNEGIGHPC 389

RESULT 9
A54411
C:Species: Bos primigenius taurus (cattle)
C:Date: 09-Sep-1994 #sequence_revision 06-Feb-1995 #text_change 18-Feb-2000
C:Accession: A54411; B38081; A48242; S65408
R;Mu., D.; Medzhradszky, K.F.; Adams, G.W.; Mayer, P.; Hines, W.M.; Burlingame, A.L.; Smi
J. Biol. Chem. 269, 9326-9332, 1994
A>Title: Primary structures for a mammalian cellular and serum copper amine oxidase.
A:Reference number: A54411; MUID:94193686; PMID:8144587
A:Accession: A54411
A:Molecule type: mRNA
A:Residues: 1-762 <MUA>
A:Cross-references: GB:S69583; NID:9546215; PIDN:AAB30397.1; PID:9546216
R;Mu., D.; Janes, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Klimman, J.P.
J. Biol. Chem. 267, 7979-7982, 1992
A>Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine o
A:Reference number: A38081; MUID:92235001; PMID:1569055
A:Accession: B38081
A:Molecule type: Protein
A:Residues: 463-465,'D',467-473,'X',475-485 <MU2>
R;Janes, S.M.; Mu., D.; Memmer, D.; Smith, A.J.; Kaur, S.; Maltby, D.; Burlingame, A.L.; f
Science 248, 981-987, 1990
A>Title: A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of
A:Reference number: A48242; MUID:90260648; PMID:2111581
A:Accession: A48242
A>Status: preliminary
A:Molecule type: Protein
A:Residues: 468-472 <UAN>
R;De Biase, D.; Agostinelli, E.; de Matteis, G.; Mondovi, B.; Morpurgo, L.
Eur. J. Biochem. 237, 93-99, 1996
A>Title: Half-of-the-sites reactivity of bovine serum amine oxidase. Reactivity and chem
A:Reference number: S65408; MUID:96203913; PMID:8620899
A:Accession: S65408
A:Molecule type: Protein
A:Residues: 463-469,'X',471-487 <DEB>
C:Superfamily: amiloride-binding protein
C:Keywords: copper; glycoprotein; oxidoreductase; quinoxaline; topaquinine
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-762/Product: amine oxidase (copper-containing) #status predicted <MAT>
F:136,231,685/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:470/Modified site: topaquione (Iyr) #status experimental

Query Match           14.7%; Score 67; DB 2; Length 762;
Best Local Similarity 38.6%; Pred. No. 30;
Matches    22; Conservative      3; Mismatches    18; Indels    14; Gaps     3;

QY       22 IWTGL-----GEQGEGIGPEQAASPTPCASRWPSASRWPWSAGLTVDRPQL 71
          |||::|||::|||::|||::|||::|||::|||::|||::|||
DB         9 IWLTLVWGREGGVSGEGRKQCCHPSLPKCRSP---SDQPW----THPDOSQL 58

RESULT 10
T00027
brain-specific angiogenesis inhibitor 2 - human
N:Alternate names: BA12 protein
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00027
R;Shiratsuchi, T.; Nishimori, H.; Ichise, H.; Nakamura, Y.; Tokino, T.
Cyto genet. Cell Genet. 79, 103-108, 1997
A>Title: Cloning and characterization of BA12 and BA13, novel genes homologous to brain-
A:Reference number: Z14086; MUID:98194217; PMID:9533023
A:Accession: T00027
A>Status: translated from GB/EMBL/DBJ

```

A:Molecule type: mRNA

A:Residues: 1-1572 <SHI>

A:Cross-references: EMBL:AB005298; NID:g3021698; PIDN:BAA25362.1; PID:g3021699

A:Experimental source: brain

C:Genetics:

A:Gene: GDB:BAI2

A:Cross-references: GDB:9938089; OMIM:602683

A:Map position: lp35-tp35

Query Match 14.7%; Score 67; DB 2; Length 1572;
Best Local Similarity 23.8%; Pred. No. 59;
Matches 36; Conservative 3; Mismatches 26; Indels 86; Gaps 9;

QY 2 CGRRPVVAG--CGFADAHWTGLTGLGEGGGTG-----PEGQ 39
Db 196 CGR-----AAGRACGFAQ-----PGCPCGAGAGSTTTSPGPPAAHTLSNALVPGGP 244

QY 40 ASPT-----PDCASRWPRS-----ASRW- 57
Db 245 APPAEDLHGSSNDLFTTMYRGEEPEEFVKYQWPRSADEPGLYMAQTGDPAAEWS 304

QY 58 PWSA-----GLTVRDR-----PQGLGCMG 77
Db 305 PWSVGLTCGQGLQVTRSCVSPYGTLCSG 335

RESULT 11

T13951

DNA topoisomerase (EC 5.99.1.2) III - mouse

C:Species: Mus musculus (house mouse)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C:Accession: T13951

R:Seki, T.; Seki, M.; Katada, T.; Enomoto, T.

Biochim. Biophys. Acta 1396, 127-131, 1998

A:Title: Isolation of a cDNA encoding mouse DNA topoisomerase III which is highly expres

A:Reference number: Z17829; MUID:98201702; PMID:9540825

A:Accession: T13951

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1003 <SEK>

A:Cross-references: EMBL:AB006074; NID:g3061307; PIDN:BAA25662.1; PID:g3061308

C:Genetics:

A:Gene: mTOP3

C:Keywords: isomerase

Query Match 14.5%; Score 66; DB 2; Length 1003;
Best Local Similarity 28.2%; Pred. No. 49;
Matches 24; Conservative 8; Mismatches 23; Indels 30; Gaps 5;

QY 2 CGRRPVVAGC-----GFADAHWTGLTGLGEGGGTGPEGQASPTPDCAARW 50
Db 924 CAKPREQCGFFQWVDENVAPGSAAPWPG---GRGKAQR-----PEAASKR 968

QY 51 PRSASRWPSAGLTVRDRPQGLGELC 75

Db 969 PRAGSS----DAGSTVK-KPRKCSLC 989

RESULT 12

S43275

hypothetical protein 2 - Neurospora crassa retrotransposon Tad1-1

C:Species: Neurospora crassa

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 01-Dec-2000

C:Accession: S43275; S43277

R:Cambareri, E.B.; Helber, J.; Kinsey, J.A.

Mol. Gen. Genet. 242, 658-665, 1994

A:Title: Tad1-1, an active LINE-like element of Neurospora crassa.

A:Reference number: S43274; MUID:94203179; PMID:7512193

A:Accession: S43275

A:Molecule type: DNA

A:Residues: 1-1154 <CAM>

A:Cross-references: EMBL:L25662; NID:g409759; PIDN:AAA21781.1; PID:g409761

A:Accession: S43277

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-710, 'E', 712-1154 <CA2>

A:Cross-references: EMBL:L25663; NID:g409762; PIDN:AAA21792.1; PID:g409764

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993

C:Genetics:

A:Mobile element: retrotransposon Tad1-1

Query Match 14.5%; Score 66; DB 2; Length 1154;
Best Local Similarity 28.4%; Pred. No. 56;
Matches 19; Conservative 1; Mismatches 37; Indels 10; Gaps 2;

QY 3 GRPRRVVAGCGFADAHWTGLWT-----GLGEGGGTGPEGQASPTPDCAAR--WPR 52
Db 910 GRPPADKPESGAARWAEAWAPTEKLLAMAGPPESGTPPPASQIPEPTAADREWLE 969

QY 53 SASRWPW 59
Db 970 DATNREW 976

RESULT 13

AG1940

hypothetical protein alr1074 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AG1940

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AG1940

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-219 <KUR>

A:Cross-references: GB:BA000019; PIDN:BAE73031.1; PID:g17130420; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr1074

C:Superfamily: Synechocystis hypothetical protein slr0885

Query Match 14.4%; Score 65.5; DB 2; Length 219;
Best Local Similarity 31.7%; Pred. No. 13;
Matches 19; Conservative 7; Mismatches 23; Indels 11; Gaps 4;

QY 21 GLW-----TGLGEGGGTGPEGQASPTPDCAARWPSASRWPSAGLTVRDRPQGLGELC 75
Db 89 GIWTKDGTGTVGSGSGGFTLPNGAVRSPDAA--WIKKA-RW---EAIPEQKRKPAPIC 142

RESULT 14

T30630

hypothetical protein 28L - Molluscum contagiosum virus 1

N:Alternate names: MC028L

C:Species: Molluscum contagiosum virus 1

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-May-2000

C:Accession: T30630

R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.

Science 273, 813-816, 1996

A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re

A:Reference number: Z20876; MUID:96325459; PMID:8670425

A:Accession: T30630

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-268 <SEN>

A:Cross-references: EMBL:U60315; PIDN:AAC55156.1

C:Genetics:

A:Note: MC028L

Query Match 14.4%; Score 65.5; DB 2; Length 268;
Best Local Similarity 38.3%; Pred. No. 16;

Matches 18; Conservative 6; Mismatches 18; Indels 5; Gaps 2;

QY 25 GLGEGQGGI---GPEGQASPTDCASR-WPRSASRWPWSAGLTVR 66
 DB 27 GAGSASEDDAEPEPGPEFPEPAARSFRAARLPGAGGLPRR 73

RESULT 15

A24925
 beta-galactosidase (EC 3.2.1.23) - Klebsiella pneumoniae
 C:Species: Klebsiella pneumoniae
 C:Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 22-Jun-1999
 C:Accession: A24925
 J. Bacteriol. 163, 850-857, 1985
 R. Buvinger, W.E.; Riley, M.
 A:Title: Nucleotide sequence of Klebsiella pneumoniae lac genes.
 A:Reference number: A91803; MUID:85289025; PMID:3897196
 A:Accession: A24925
 A:Molecule type: DNA
 A:Residues: 1-1034 <BUV>
 A:Cross-references: GB:M11441; GB:M11416; NID:g149216; PIDN:AAA25082.1; PID:g149218
 C:Genetics:
 A:Gene: lacZ
 C:Superfamily: beta-galactosidase
 C:Keywords: glycosidase; hydrolase

Query Match 14.4%; Score 65.5; DB 2; Length 1034;
 Best Local Similarity 32.7%; Pred. No. 57;
 Matches 17; Conservative 6; Mismatches 20; Indels 9; Gaps 3;

QY 18 HW-----TGLWTGLGEGQGGIGPEGQASPTDCASRWPWSAGL 63
 DB 983 HWHKMQAEDGVWITL-DGLHGVG--GDSWTFVLPWLLSQTRWQYEVSL 1031

Search completed: August 13, 2004, 09:07:51
 Job time : 18 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2004, 08:56:32 ; Search time 13 Seconds

(without alignments)
316.426 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456

Sequence: 1 MCGRRRVRSAGCGFADAHWT.....SAGLTVRDRPQIGELCMGRG 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	17.8	367	1 LHX8 MOUSE	Q35652 mus musculus
2	73.5	16.1	1729	1 TABP_HUMAN	Q9C0C2 homo sapien
3	70	15.4	1669	1 C314_HUMAN	P02462 homo sapien
4	68.5	15.0	310	1 NF1U_AZOBK	Q43909 azospirillum
5	67.5	14.8	1733	1 VNUA_PVKKA	P23485 pseudorabies
6	67	14.7	333	1 NK32_HUMAN	P78367 homo sapien
7	67	14.7	547	1 RM56_HUMAN	P83111 homo sapien
8	67	14.7	762	1 AOCX_BOVIN	Q29437 bos taurus
9	67	14.7	762	1 AOCY_BOVIN	O46406 bos taurus
10	67	14.7	1572	1 BA12_HUMAN	O62241 homo sapien
11	66	14.5	1003	1 TF3A_MOUSE	O70157 mus musculus
12	65.5	14.4	1034	1 BGAL_KLEPN	P06219 klebsiella
13	64.5	14.1	1669	1 CA14_MOUSE	P02463 mus musculus
14	64	14.0	576	1 Z384_HUMAN	Q8tf68 homo sapien
15	63.5	13.9	275	1 TRVA_HUMAN	P15157 homo sapien
16	63.5	13.9	596	1 FIBA_BOVIN	P02672 bos taurus
17	63	13.8	376	1 FRAE_EORPE	Q00879 bordetella
18	63	13.8	579	1 Z384_RAT	Q06q14 rattus norv
19	63	13.8	1216	1 AEGP_RAT	Q63191 rattus norv
20	63	13.8	1838	1 CHA2_HUMAN	P20908 homo sapien
21	62	13.6	132	1 CHA2_BOVMO	P08825 bombyx mori
22	62	13.6	426	1 EGL1_HUMAN	Q9gz19 homo sapien
23	62	13.6	485	1 ONC2_HUMAN	O95948 homo sapien
24	61.5	13.5	745	1 ATCS_SYNF3	P73241 synchocyst
25	61	13.4	651	1 CSP6_HUMAN	Q9nv66 h cofactor
26	61	13.4	1139	1 M2A2_HUMAN	P49641 homo sapien
27	61	13.4	1171	1 GLG1_RAT	Q62638 rattus norv
28	60.5	13.3	164	1 LWA_ACTEO	Q16998 actinia equ
29	60.5	13.3	389	1 WN1E_MOUSE	P48614 mus musculus
30	60.5	13.3	520	1 CMCN_NOCLA	Q51080 noxia la
31	60	13.2	161	1 CHB2_BOVMO	P08828 bombyx mori
32	60	13.2	226	1 NUKM_NEUCR	O47950 neurospora
33	60	13.2	616	1 REF5_HUMAN	P48382 homo sapien

ALIGNMENTS

RESULT 1

LHX8 MOUSE	ID	LHX8_MOUSE	STANDARD;	PRT;	367 AA.
AC	O35652	070163	089707;		
DT	30-MAY-2000	(Rel. 39, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	15-MAR-2004	(Rel. 43, Last annotation update)			
DE	LHX8/homeobox protein Lhx8 (LIM homeodomain Lhx7) (L3).				
GN	LHX8 OR LHX7				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Chordata; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ICR; TISSUE=Brain;				
RX	MEDLINE=97083417; PubMed=8929991;				
RA	Matsumoto K., Tanaka T., Furiyama T., Kashiwara Y., Mori T., Ishii N.,				
RA	Kitanaka J., Takemura M., Tohyama M., Wanaoka A.;				
RT	"L3, a novel murine LIM-homeodomain transcription factor expressed in				
RT	the ventral telencephalon and the mesenchyme surrounding the oral				
RT	cavity.";				
RL	Neurosci. Lett. 204:113-116(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=129/SVJ; TISSUE=Liver;				
RX	MEDLINE=98260681; PubMed=9598319;				
RA	Kitanaka J.-I., Takemura M., Matsumoto K., Mori T., Wanaoka A.;				
RT	"Structure and chromosomal localization of a murine LIM/homeobox gene,				
RT	Lhx8.";				
RL	Genomics 49:307-309(1998).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	MEDLINE=98237706; PubMed=9570771;				
RA	Grigoriou M., Tucker A.S., Sharpe P.T., Pachnis V.;				
RT	"Expression and regulation of Lhx6 and Lhx7, a novel subfamily of LIM				
RT	homeodomain encoding genes, suggests a role in mammalian head				
RT	development.";				
RL	Development 125:2063-2074(1998).				
CC	-I- FUNCTION: Transcription factor involved in differentiation of				
CC	certain neurons and mesenchymal cells.				
CC	-I- SUBCELLULAR LOCATION: Nuclear (Probable).				
CC	-I- DEVELOPMENTAL STAGE: Limited spatially to the medical ganglionic				
CC	emience and the mesenchyme surrounding the oral cavity and				
CC	temporally from middle embryonic to early postnatal development.				
CC	-I- SIMILARITY: Contains 1 homeobox domain.				
CC	-I- SIMILARITY: Contains 2 LIM zinc-binding domains.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				

P39881 canis famill
Q95076 homo sapien
Q9b2z6 homo sapien
P06475 herpes simp
Q89730 herpes simp
P08933 sus scrofa
Q9xsm2 ovis aries
P08127 pseudomonas
P06622 pseudomonas
P27887 pseudomonas
P23490 homo sapien
P38059 pseudomonas

34 60 13.2 975 1 CUTI_CANFA
35 59.5 13.0 343 1 ALX3_HUMAN
36 59.5 13.0 473 1 RT4R_HUMAN
37 59.5 13.0 480 1 VGLC_HSV23
38 59.5 13.0 480 1 VGLC_HSV2H
39 59.5 13.0 926 1 PERP_PIG
40 59 12.9 273 1 TRY1_SHEEP
41 59 12.9 307 1 NAHH_PSEPU
42 59 12.9 307 1 XYLE_PSEPU
43 59 12.9 307 1 XYLE_PSEAE
44 59 12.9 316 1 LORI_HUMAN
45 59 12.9 384 1 CARA_PSEST

EMBL; D49658; BAA21649.1; ALT_INIT.
DR EMBL; AB007596; BAA28628.1; -;
DR EMBL; AB007588; BAA28628.1; JOINED.
DR EMBL; AB007589; BAA28628.1; JOINED.
DR EMBL; AB007590; BAA28628.1; JOINED.
DR EMBL; AB007591; BAA28628.1; JOINED.
DR EMBL; AB007592; BAA28628.1; JOINED.
DR EMBL; AB007593; BAA28628.1; JOINED.
DR EMBL; AB007594; BAA28628.1; JOINED.
DR EMBL; AB007595; BAA28628.1; JOINED.
DR EMBL; AB007596; BAA28628.1; JOINED.
DR HSSP; P06601; 1FJL.
DR TRANSFAC; T04185; -;
DR TRANSFAC; T04191; -;
DR MGI; 1096343; Lhx8.
DR GO; GO:0042475; P:ontogenesis (sensu Vertebrata); IMP.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR001781; LIM.
DR InterPro; IPR007107; LIM homeo.
DR Pfam; PF00046; homeobox_1.
DR Pfam; PF00412; LIM; 2.
DR ProDom; PD000010; Homeobox; 1.
DR ProDom; PD000094; LIM; 2.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00132; LIM; 2.
DR PROSITE; PS00478; LIM DOMAIN 1; 2.
DR PROSITE; PS50023; LIM DOMAIN 2; 2.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS50071; HOMEBOX 2; 1.
KW Homeobox; DNA-binding; Nuclear protein; Repeat; LIM domain;
KW Metal-binding; Zinc; Transcription regulation.
FT DOMAIN 96
FT DOMAIN 157
FT DOMAIN 210
FT DNA_BIND 246
FT CONFLICT 18
FT CONFLICT 27
FT CONFLICT 42
FT CONFLICT 51
FT CONFLICT 72
FT CONFLICT 87
FT CONFLICT 327
FT CONFLICT 337
FT CONFLICT 341
FT CONFLICT 347
FT CONFLICT 351
SQ SEQUENCE 367 AA; 40696 MW; 4E104F53A008A31F CRC64;
Query Match 17.8%; Score 81; DB 1; Length 367;
Best Local Similarity 29.3%; Pred No. 0.51; Indels 30; Gaps 4;
Matches 27; Conservative 10; Mismatches 10; RTRKGAGEGLVNPGE-AGDEDCSSSGPLSPSSPQSM 83
QY 2 CGRPRVSGCGFADAHWTGLTGLGEGGEGIGEGQASPTPCASRWPSASRWFW-- 59
Db 36 CGRPAALAG-----RTRKGAGEGLVNPGE-AGDEDCSSSGPLSPSSPQSM 83
QY 60 -----SAGLVTRDR--POLGELC 75
Db 84 ASGSVCPGKVCSSCGLEIVDKYLLKNDLC 115
RESULT 2
TAPP HUMAN
ID TAPP HUMAN STANDARD; PRT; 1729 AA.
AC Q9C0C2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 182 kDa tankyrase 1-binding protein.
GN TNKS1BP1 OR TAB182 OR KIAA1741.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1] SEQUENCE FROM N.A.
RP TISSUE=Placenta, and Testis;
RX MEDLINE=21950796; PubMed=11854288;
RA Seimiya H., Smith S.;
RT "The telomeric poly(ADP-ribose) polymerase, tankyrase 1, contains
multiple binding sites for telomeric repeat binding factor 1 (TRF1)
and a novel acceptor, 182-kDa tankyrase-binding protein (TAB182).";
RL J. Biol. Chem. 277:14116-14126(2002).
[2] SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
[3] SEQUENCE OF 495-1729 FROM N.A.
RP TISSUE=Spleen;
RX MEDLINE=22579292; PubMed=12693554;
RA Jikuya H., Takano J., Kikuno R., Hirose M., Nagase T., Nomura N.,
Ohara O.;
RT "Characterization of long cDNA clones from human adult spleen. II. The
complete sequences of 81 cDNA clones.";
RL DNA Res. 10:49-57(2003).
CC -!- SUBUNIT: Binds to the ANK repeat domain of TNKS1 and TNKS2.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. Colocalizes with
chromosomes during mitosis, and in the cytoplasm with cortical
actin. SPECIFICITY: Detected in testis, ovary, lung, skeletal
muscle, heart, prostate and pancreas, and at very low levels in
brain and peripheral blood leukocytes.
CC -!- PTM: ADP-ribosylated by TNKS1 (in vitro).
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
in position 1071, 1097 and 1457.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

DR EMBL; AF441771; AAM15531.1; -;
DR EMBL; AB051528; BAB21832.2; ALT_INIT.
DR EMBL; AK074113; BAB84939.1; ALT_FRAME.
DR Genew; HGNC:19081; TNKS1BP1.
DR GK; Q9C0C2; -;
DR MIM; 607104; -;
DR GO; GO:0005737; C:cytoplasm; NAS.
DR GO; GO:0005724; C:nuclear telomeric heterochromatin; NAS.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0030506; F:ankyrin binding; NAS.
DR GO; GO:0019899; F:enzyme binding; NAS.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
DR InterPro; IPR008979; Gal_bind_like.
KW Nuclear protein; Chromosomal protein; ADP-ribosylation.
FT DOMAIN 2
FT DOMAIN 103
FT DOMAIN 127
FT DOMAIN 1767
FT DOMAIN 210
FT DOMAIN 1340
FT DOMAIN 1010
FT DOMAIN 1542
FT DOMAIN 1450
FT DOMAIN 1572
FT DOMAIN 1629
FT DOMAIN 1635
FT DOMAIN 1729
FT DOMAIN 1733
FT CONFLICT 84
FT CONFLICT 322
FT CONFLICT 388
FT CONFLICT 388
FT CONFLICT 554
FT CONFLICT 554
FT CONFLICT 604
FT CONFLICT 604
P -> L (IN REF. 2).
S -> T (IN REF. 2).
S -> P (IN REF. 2).
Q -> H (IN REF. 3).
P -> S (IN REF. 1).

FT CONFLICT 1450 1450 F -> S (IN REF. 2).
 SQ SEQUENCE 1729 AA; 181814 MW; C65F38FA37045C4A CRC64;
 Query Match 16.1%; Score 73.5; DB 1; Length 1729;
 Best Local Similarity 27.6%; Pred. No. 12;
 Matches 24; Conservative 4; Mismatches 30; Indels 29; Gaps 2;
 QY 6 RVVSGCGFADAHWTGLTGLGEGGEGGIG-----PEGQASPTPD 45
 DB 1470 RESAASGLG-----GLLEEGCAGAGAAQEEVLEPGRDSPSPWRPQDGEASQTED 1520
 QY 46 CASRWPRGASRWPSAGLTVDRPQLG 72
 DB 1521 VDGITGSSAARWSDGPAQTSRRPSQG 1547
 RESULT 3
 ID CA14 HUMAN STANDARD; PRT; 1669 AA.
 AC P02462;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89340433; PubMed=2701944;
 RA Soininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;
 RT "Structural organization of the gene for the alpha 1 chain of human
 type IV collagen.";
 RL J. Biol. Chem. 264:13565-13571(1989).
 RN [2]
 RP SEQUENCE OF 46-1257 FROM N.A.
 RX TISSUE=Placenta;
 RC MEDLINE=88083584; PubMed=3691802;
 RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
 RT "Complete primary structure of the alpha 1-chain of human basement
 membrane (type IV) collagen.";
 RL FEBS Lett. 225:188-194(1987).
 RN [3]
 RP SEQUENCE OF 1-943 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88029471; PubMed=3311751;
 RA Brazel D., Oberbauer I., Dieringer H., Babel W., Glanville R.W.,
 RT Deutzmann R., Kuehn K.;
 RT "Completion of the amino acid sequence of the alpha 1 chain of human
 basement membrane collagen (type IV) reveals 21 non-triplet
 interruptions located within the collagenous domain.";
 RL Eur. J. Biochem. 168:529-536(1987).
 RN [4]
 RP SEQUENCE OF 28-243.
 RX MEDLINE=86004708; PubMed=403082;
 RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
 RT "Amino acid sequence of the N-terminal aggregation and cross-linking
 region (7S domain) of the alpha 1 (IV) chain of human basement
 membrane collagen.";
 RL Eur. J. Biochem. 152:213-219(1985).
 RN [5]
 RP SEQUENCE OF 534-1447.
 RX MEDLINE=85003629; PubMed=6434307;
 RA Babel W., Glanville R.W.;
 RT "Structure of human basement-membrane (type IV) collagen. Complete
 amino-acid sequence of a 914-residue-long pepsin fragment from the
 alpha 1(IV) chain.";
 RL Eur. J. Biochem. 143:545-556(1984).
 RN [6]
 RP SEQUENCE OF 1256-1669 FROM N.A.
 RX MEDLINE=85207819; PubMed=2581969;

RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
 RT Cheung M.-C., Prockop D.J., Boyd C.D.;
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
 procollagen reveal an unusual homology of amino acid sequences in two
 halves of the carboxyl-terminal domain.";
 RL J. Biol. Chem. 260:7681-7687(1985).
 RN [7]
 RP SEQUENCE OF 1259-1669 FROM N.A.
 RX MEDLINE=85216555; PubMed=2582422;
 RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
 RT Kefalides N.A., Myers J.C.;
 RT "Restricted homology between human alpha 1 type IV and other
 procollagen chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
 RN [8]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=89034231; PubMed=3182844;
 RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 collagen are divergently encoded on opposite DNA strands and have an
 overlapping promoter region.";
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN [9]
 RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
 RC TISSUE=Placenta;
 RX MEDLINE=89005112; PubMed=2844531;
 RA Siebold B., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 carboxyterminal, non-collagenous aggregation and cross-linking domain
 of basement-membrane type IV collagen.";
 RL Eur. J. Biochem. 176:617-624(1988).
 CC -I- FUNCTION: Type IV collagen is the major structural component of
 glomerular basement membranes (GBM), forming a 'chicken-wire'
 meshwork together with laminins, proteoglycans and entactin/
 nidogen.
 CC -I- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
 alpha 6(IV), each of which can form a triple helix structure
 with 2 other chains to generate type IV collagen network.
 CC -I- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 domain (NC1) at their C-terminus, frequent interruptions of the
 G-X-Y repeats in the long central triple-helical domain (which may
 cause flexibility in the triple helix), and a short N-terminal
 triple-helical 7S domain.
 CC -I- PTM: Lysines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
 CC -I- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -I- PTM: Type IV collagens contain numerous cysteine residues which
 are involved in inter- and intramolecular disulfide bonding. 12 of
 these, located in the NC1 domain, are conserved in all known type
 IV collagens.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M26576; AAA53098.1; JOINED.
 DR EMBL; J04217; AAA53098.1; JOINED.
 DR EMBL; M26550; AAA53098.1; JOINED.
 DR EMBL; M26540; AAA53098.1; JOINED.
 DR EMBL; M26542; AAA53098.1; JOINED.
 DR EMBL; M26543; AAA53098.1; JOINED.
 DR EMBL; M26544; AAA53098.1; JOINED.
 DR EMBL; M26545; AAA53098.1; JOINED.
 DR EMBL; M26546; AAA53098.1; JOINED.
 DR EMBL; M26547; AAA53098.1; JOINED.
 DR EMBL; M26537; AAA53098.1; JOINED.
 DR EMBL; M26538; AAA53098.1; JOINED.
 DR EMBL; M26548; AAA53098.1; JOINED.

DR EMBL; M26549; AAA53098.1; JOINED.
DR EMBL; M26551; AAA53098.1; JOINED.
DR EMBL; M26552; AAA53098.1; JOINED.
DR EMBL; M26553; AAA53098.1; JOINED.
DR EMBL; M26554; AAA53098.1; JOINED.
DR EMBL; M26555; AAA53098.1; JOINED.
DR EMBL; M26556; AAA53098.1; JOINED.
DR EMBL; M26557; AAA53098.1; JOINED.
DR EMBL; M26558; AAA53098.1; JOINED.
DR EMBL; M26559; AAA53098.1; JOINED.
DR EMBL; M26560; AAA53098.1; JOINED.
DR EMBL; M26561; AAA53098.1; JOINED.
DR EMBL; M26562; AAA53098.1; JOINED.
DR EMBL; M26563; AAA53098.1; JOINED.
DR EMBL; M26564; AAA53098.1; JOINED.
DR EMBL; M26565; AAA53098.1; JOINED.
DR EMBL; M26566; AAA53098.1; JOINED.
DR EMBL; M26567; AAA53098.1; JOINED.
DR EMBL; M26568; AAA53098.1; JOINED.
DR EMBL; M26569; AAA53098.1; JOINED.
DR EMBL; M26570; AAA53098.1; JOINED.
DR EMBL; M26571; AAA53098.1; JOINED.
DR EMBL; M26572; AAA53098.1; JOINED.
DR EMBL; M26573; AAA53098.1; JOINED.
DR EMBL; M26574; AAA53098.1; JOINED.
DR EMBL; M26575; AAA53098.1; JOINED.
DR EMBL; Y00706; CAA29075.1; -.
DR EMBL; X05561; CAA29075.1; -.
DR EMBL; M10940; AAA52006.1; -.
DR EMBL; M11315; AAA52042.1; -.
DR PIR; S16876; CGHU4B.
DR Genew; HGNC:2202; COL4A1.
DR MIM; 120130; -.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g helix; 6.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NON-HELICAL REGION (NC1).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).
FT DISULFID 1460 1551 OR 1548.
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1548.
FT DISULFID 1570 1665 OR 1662.
FT DISULFID 1604 1662 OR 1665.
FT DISULFID 1616 1622
FT CONFLICT 227 328 SG -> KE (IN REF. 4).
FT CONFLICT 241 241 G -> K (IN REF. 4).
FT CONFLICT 319 319 N -> D (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 944 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 S -> K (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
FT CONFLICT 1669 AA; 160611 MW; 3BBA6DFFB9B8A84 CRC64;
SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFFB9B8A84 CRC64;

Query Match 15.4%; Score 70; DB 1; Length 1669;
Best Local Similarity 31.1%; Pred. NO. 25;
Matches 23; Conservative 7; Mismatches 24; Indels 20; Gaps 3;

QY 10 AGCGFADAHWT-----GLWTGLG-----EGQEGGIGPEGQASPTPD-----CASR 49
Db 34 SCGKCDCHGVKGQKGERGLPGLQGVIFGPMQGPQGGPGKGDIGELPCTKCTR 93
QY 50 WPRSASRWPWSAGL 63
Db 94 GPPGASGYGNGPL 107

RESULT 4
NIFU AZOBR
ID NIFU AZOBR STANDARD; PRT; 310 AA.
AC Q4309; P70726;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Nitrogen fixation protein nifu.
GN NIFU.
OS Azospirillum brasilense.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Azospirillum.
OX NCBI_TaxID=192;
RN [1]
RP SEQUENCE FROM N.A.
RA Frazzon J.S., Schrank I.S.;
RT "The nifu gene from Azospirillum brasilense.";
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98164354; PubMed=9503607;
RA Frazzon J.S., Schrank I.S.;
RT "Sequencing and complementation analysis of the nifUSV genes from
Azospirillum brasilense.";
RL FEMS Microbiol. Lett. 159:151-157(1998).
CC -!- FUNCTION: MAY BE INVOLVED IN THE FORMATION OR REPAIR OF [FE-S]
CLUSTERS PRESENT IN IRON-SULFUR PROTEINS (POTENTIAL).
CC -!- SIMILARITY: Belongs to the nifu family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; L34354; AAA22184.1; -.
CC EMBL; U26437; AAC46176.1; -.
DR InterPro; IPR007419; Fer2_BFD.
DR InterPro; IPR001075; Nifu_C.
DR InterPro; IPR002871; Nifu_N.
DR Pfam; PF04324; fer2_BFD; 1.
DR Pfam; PF01106; Nifu-like; 1.
DR Pfam; PF01592; Nifu_N; 1.
DR ProDom; PD002830; Nifu_C; 1.
KW Nitrogen fixation.
FT CONFLICT 56 56 N -> K (IN REF. 1).
FT CONFLICT 133 136 ESSE -> AAGT (IN REF. 1).
FT CONFLICT 141 144 CFGI -> DRRQ (IN REF. 1).
FT CONFLICT 149 152 IERA -> DRRR (IN REF. 1).
FT CONFLICT 201 202 MISSING (IN REF. 1).
FT CONFLICT 212 223 TPSSRRPRRH -> IAYQAGPKAEA (IN REF. 1).
FT CONFLICT 230 231 NV -> TC (IN REF. 1).
FT CONFLICT 235 236 OK -> RR (IN REF. 1).
FT CONFLICT 243 244 DW -> EL (IN REF. 1).
FT CONFLICT 281 282 GT -> AS (IN REF. 1).
FT CONFLICT 303 307 ASLLP -> LSA (IN REF. 1).
SQ SEQUENCE 310 AA; 33201 MW; EF9F474F91255FEF CRC64;


```

Query Match          15.0%; Score 68.5; DB 1; Length 310;
Best Local Similarity 51.9%; Pred.No.7.3;
Matches 14; Conservative 1; Mismatches 11; Indels 1; Gaps 1;

QY      32   GGIGECQAQPPTPCASRWPRSRWP 58
DB      193   GAVGPAQAPSPTPPARSGWTPS-SRWP 218
           :|||
           |||||
           |||||

RESULT 5
VNUA PRVKA
ID VNUA PRVKA STANDARD; PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable nuclear antigen.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; PubMed=2171211;
RA Vitek C., Kozmik Z., Faces V., Schirm S., Schwytzer M.;
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT oriented open reading frame: characterization of their promoter and
RT enhancer regions.";
RL Virology 179:365-377(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M34651; AAA47471.1; --
DR PIR; B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475BB5E2 CRC64;

Query Match          14.8%; Score 67.5; DB 1; Length 1733;
Best Local Similarity 33.8%; Pred.No.46;
Matches 25; Conservative 3; Mismatches 39; Indels 7; Gaps 2;

QY      8   VSAGCGFADAHWTGLWGTCGGQGEGGTGPQGASPTDPCASRWP--RSASRWPSAGLTV 65
DB      967   VAGGAGEA-----GLGNAGALGAGALGACGAGCGPGAGEAGGARRRRRWDEAGLLG 1021
           :|||
           |||||
           |||||

QY      66   RDRPQLGELCMGRG 79
DB      1022   PERQAGRGLRGPG 1035
           :|||
           |||||
           |||||

RESULT 6
NK32_HUMAN
ID NK32_HUMAN STANDARD; PRT; 333 AA.
AC P78367;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein Nkx-3.2 (Bagpipe homeobox protein homolog 1).
DE BAPX1 OR NKX3B.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

RESULT 7
RM56 HUMAN
ID RM56 HUMAN STANDARD; PRT; 547 AA.
AC P83111; P83096;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mitochondrial 39S ribosomal protein L56 (MRP-L56) (Serine beta
DE lactamase-like protein LACTB).
GN LACTB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Li W.B., Gruber C., Jessee J., Polayes D.;
RT "Full-length cDNA libraries and normalization.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Birren B., Linton L., Nusbaum C., Lander E.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-373 FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP IDENTIFICATION, AND CONCEPTUAL TRANSLATION.
RA Koc E.C., Burkhardt W., Blackburn K., Schlatter D.M., Moseley A.,
RA Sprengelli L.L.;
RT "The large subunit of the mammalian mitochondrial ribosome:
RT Identification of the full complement of ribosomal proteins present.";
RL Submitted (AUG-2001) to Swiss-Prot.
RN [5]
RP IDENTIFICATION, AND CONCEPTUAL TRANSLATION.
RA MEDLINE=21564197; PubMed=11707067;
RA Smith T.S., Southan C., Ellington K., Campbell D., Tew D.G.,
RA Debouck C.;
RT "Identification, genomic organization, and mRNA expression of LACTB,
RT encoding a serine beta-lactamase-like protein with an amino-terminal
RT transmembrane domain.";
RL Genomics 78:12-14(2001).
CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed predominantly in skeletal
CC muscle.
CC -!- SIMILARITY: BELONGS TO THE L56E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL553126; -; NOT ANNOTATED_CDS.
CC EMBL; AL553134; -; NOT ANNOTATED_CDS.
CC EMBL; AC026817; -; NOT ANNOTATED_CDS.
CC EMBL; AK027808; BAB5384.1; ALT_TERM.
CC EMBL; HGNC:16468; LACTB.
CC InterPro; IPR001466; Beta_lactamase.
CC
DR EMBL; AL553126; -; NOT ANNOTATED_CDS.
DR EMBL; AL553134; -; NOT ANNOTATED_CDS.
DR EMBL; AC026817; -; NOT ANNOTATED_CDS.
DR EMBL; AK027808; BAB5384.1; ALT_TERM.
DR EMBL; HGNC:16468; LACTB.
DR InterPro; IPR001466; Beta_lactamase.
DR

```

```

DR Pfam; PF00144; beta-lactamase; 1.
KW Ribosomal protein; Mitochondrion.
SQ SEQUENCE 547 AA; 60693 MW; C4BDE6B8EF39169A CRC64;
Query Match 14.7%; Score 67; DB 1; Length 547;
Best Local Similarity 24.4%; Pred. No. 18;
Matches 19; Conservative 10; Mismatches 27; Indels 22; Gaps 2;
QY 5 PRVRSAGCGFADAHWTGLTGLTGLGEGEGIG-----PEGASPTP 44
Db 15 PRLASSCGRCGVHORAGLPLPLGHWGVLGLGLGLGVLGVLGRLGAPAAQSPADDP 74
QY 45 DCA--SRWPRSRASRWPS 60
Db 75 EASPLAEPPEQOSLAPWS 92
RESULT 8
AOCX_BOVIN STANDARD; PRT; 762 AA.
ID AOCX_BOVIN
AC Q29437;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Copper amine oxidase, liver isozyme precursor (EC 1.4.3.6) (Amine
DE oxidase [copper-containing]) (Serum amine oxidase) (SAO).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=94193686; PubMed=8144587;
RA Mu D., Medzihradsky K.F., Adams G.W., Mayer P., Hines W.M.,
RA Burlingame A.L., Smith A.J., Cai D., Klinman J.P.;
RT "Primary structures for a mammalian cellular and serum copper amine
RT oxidase.";
RL J. Biol. Chem. 269:9926-9932(1994).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=93090748; PubMed=1457410;
RA Jones S.M., Palic M.M., Scaman C.H., Smith A.J., Brown D.E.,
RA Dooley D.M., Mure M., Klinman J.P.;
RT "Identification of topaquinine and its consensus sequence in copper
RT amine oxidases.";
RL Biochemistry 31:12147-12154(1992).
CC -!- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -!- ENZYME REGULATION: INHIBITED BY AMILORIDE IN A COMPETITIVE MANNER.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Liver.
CC -!- PTM: Topaquinine (TPQ) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (By similarity).
CC -!- SIMILARITY: Belongs to the copper/topaquinine oxidase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S69583; AAB30397.1; -
CC EMBL; L27218; AAA30525.1; -
CC FR; A54411; A54411.
CC InterPro; IPR000269; CuNH_oxidase.
CC Pfam; PF01179; Cu_amine_oxid; 1.
CC Pfam; PF02727; Cu_amine_oxid2; 1.
CC Pfam; PF02728; Cu_amine_oxid3; 1.

```

```

DR PRINTS; PR00766; CUDAOXIDASE.
DR PROSITE; PS01164; COPPER AMINE_OXID_1; 1.
DR PROSITE; PS01165; COPPER AMINE_OXID_2; 1.
DR KW Oxidoreductase; Copper; TPQ; Glycoprotein; signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 762 COPPER AMINE OXIDASE, LUNG ISOZYME.
FT MOD_RES 470 470 TOPAQUINONE (BY SIMILARITY).
FT METAL 519 519 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT METAL 683 683 COPPER (POTENTIAL).
FT BINDING 672 672 AMILORIDE (BY SIMILARITY).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 762 AA; 84883 MW; BB43D04776744F2 CRC64;

Query Match 14.7%; Score 67; DB 1; Length 762;
Best Local Similarity 38.6%; Pred. No. 24; Gaps 14; Gaps
Matches 22; Conservative 3; Mismatches 18; Indels

QY 22 LWTGL-----GEGGEGGIGGEGGASPTTPDCASRWPSASRWPSAGLTVDRPQL 71
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 LWTLLVMGREGVSGEGVGKQCHPSLPFCPSRP---SQPW---THPDQSL 58
||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
BA12_HUMAN STANDARD; PRT; 1572 AA.
ID AC O60241;
IC BA12_HUMAN
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Brain-specific angiogenesis inhibitor 2 precursor.
DE BA12.
GN GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RA MEDLINE=98194217; PubMed=9533023;
RX Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;
RT "Cloning and characterization of BA12 and BA13, novel genes homologous
RT to brain-specific angiogenesis inhibitor 1 (BA11).";
RL Cyogenet. Cell Genet. 79:103-108(1997).
CC -!- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN
CC HEART, THYMUS, SKELETAL MUSCLE, AND DIFFERENT CELL LINES.
CC -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC -!- SIMILARITY: Contains 4 TSP type-1 domains.
CC -!- SIMILARITY: Contains 1 GPS domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL outstat
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for comm
CC entities requires a license agreement (See http://www.isb-sib.ch/anno
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB005298; BAA25362.1; -.
CC PIR; T00027; T00027.
CC Genew; HGNC; 944; BA12.
CC MIM; 602683; -.
CC InterPro; IPR000932; GPCR_secretin.
CC InterPro; IPR001879; hormn_receptor.
CC InterPro; IPR000203; PKD_cys_rich.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF00002; 7tm 2; 1.

```

DR Pfam; PF01825; GPS; 1.
 DR Pfam; PF02793; HRM; 1.
 DR Pfam; PF00090; tsp; 1; 4.
 DR SMART; SM00303; GPS; 1.
 DR SMART; SM00008; HORM; 1.
 DR SMART; SM00209; TSPL; 4.
 DR PROSITE; PS00221; GPS; 1.
 DR PROSITE; PS00649; G-PROTEIN RECF_F2_1; FALSE NEG.
 DR PROSITE; PS00650; G-PROTEIN RECF_F2_2; FALSE NEG.
 DR PROSITE; PS00227; G-PROTEIN RECF_F2_3; 1.
 DR PROSITE; PS00261; G-PROTEIN RECF_F2_4; 1.
 DR PROSITE; PS00092; TSPL; 4.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Repeat.
 FT SIGNAL. 1 20
 FT CHAIN 21 1572
 FT DOMAIN 21 924
 FT TRANSMEM 925 945
 FT DOMAIN 946 953
 FT TRANSMEM 954 974
 FT DOMAIN 975 982
 FT TRANSMEM 983 1003
 FT DOMAIN 1004 1024
 FT TRANSMEM 1025 1045
 FT DOMAIN 1046 1066
 FT TRANSMEM 1067 1087
 FT DOMAIN 1088 1141
 FT TRANSMEM 1142 1162
 FT DOMAIN 1163 1168
 FT TRANSMEM 1169 1189
 FT DOMAIN 1190 1572
 FT DOMAIN 297 350
 FT DOMAIN 352 405
 FT DOMAIN 407 460
 FT DOMAIN 463 516
 FT DOMAIN 517 559
 FT DOMAIN 559 911
 FT DOMAIN 911 122
 FT DOMAIN 122 180
 FT DOMAIN 177 180
 FT DOMAIN 222 225
 FT DOMAIN 1303 1306
 FT DOMAIN 1352 1358
 FT DOMAIN 1413 1418
 FT CARBOHYD 94 94
 FT CARBOHYD 179 179
 FT CARBOHYD 180 180
 FT CARBOHYD 344 344
 FT CARBOHYD 425 425
 FT CARBOHYD 548 548
 FT CARBOHYD 633 633
 FT CARBOHYD 855 855
 SQ SEQUENCE 1572 AA; 171140 MW; A9775645B77BC285 CIRC64;
 Query Match 14.7%; Score 67; DB 1; Length 1572;
 Best Local Similarity 23.8%; Pred. No. 47;
 Matches 36; Conservative 3; Mismatches 26; Indels 86; Gaps 9;
 QY 2 CGRRPRVSAG--CGFADAHWTGLMTGLGEGGEGIG-----PEQG 39
 DB 196 CGR-----AAGRACGFAQ-----PGCSCPGAGAGSTTTTSPGPPAAHTLSNALVPGGP 244
 QY 40 ASPT-----PDCASRMPRS-----ASRW- 57
 DB 245 APPAEADLHSSNDLFTTEMRYGEEPEEPKVTQMPRSADPEGLYMAQTGDDPAEWS 304
 QY 58 PWSA-----GLTVDRD-----POLGELCMG 77
 DB 305 PWSVCSLCGCGQLQVTRTSVSSVPGTLCSG 335
 RESULT 11
 TP3A MOUSE
 ID TP3A MOUSE PRT; 1003 AA.
 AC Q70157;

DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA topoisomerase III alpha (EC 5.99.1.2).
 GN TOP3A OR TOP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX STRAIN=BALB/c; Tissue=Testis;
 RX MEDLINE=98201702; PubMed=9540825;
 RA Seki T.; Seki M.; Katada T.; Enomoto T.;
 RT Isolation of a cDNA encoding mouse DNA topoisomerase III which is
 RL highly expressed at the mRNA level in the testis.;
 RL Biochim. Biophys. Acta 1396:127-131(1998).
 CC -!- FUNCTION: Reduces the number of supercoils in a highly negatively
 CC supercoiled DNA.
 CC -!- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
 CC DNA, followed by passage and rejoining, in testis.
 CC -!- TISSUE SPECIFICITY: Highly expressed in testis.
 CC -!- SIMILARITY: Belongs to the prokaryotic type I/III topoisomerase
 CC family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB006074; BAA25662.1; -
 CC PIR; T13951; T13951.
 CC MGD; MG1:1197527; Top3a.
 DR InterPro; IPR000380; DNA topoisomerase.
 DR InterPro; IPR003601; DNATopi_ATP_bind.
 DR InterPro; IPR003602; DNATopi_DNA_bind.
 DR InterPro; IPR006171; Toprim Dom.
 DR InterPro; IPR006154; Toprim_sub.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF01131; Topoisom_Dac; 1.
 DR Pfam; PF01751; Toprim; 1.
 DR Pfam; PF01396; zf-C4_Topoiso; 1.
 DR Pfam; PF00098; zf-CCHC; 1.
 DR PRINTS; PR00417; PRTPISMRAESI.
 DR SMART; SM00437; TOP1ac; 1.
 DR SMART; SM00436; TOP1bc; 1.
 DR SMART; SM00493; TOP1RM; 1.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
 DR PROSITE; PS00396; DNA-binding, Repeat; Zinc-finger.
 KW Isomerase; Topoisomerase; DNA-binding, Repeat; Zinc-finger.
 FT ACT_SITE 362 362 DNA CLEAVAGE (BY SIMILARITY).
 FT ZN_FING 658 685 C4-TYPE (POTENTIAL).
 FT DOMAIN 814 925 2 X 27 AA APPROXIMATE REPEATS.
 FT REPEAT 814 841 1.
 FT REPEAT 898 925 2.
 SQ SEQUENCE 1003 AA; 112358 MW; 254C738E746EB495 CRC64;
 Query Match 14.5%; Score 66; DB 1; Length 1003;
 Best Local Similarity 28.2%; Pred. No. 39;
 Matches 24; Conservative 8; Mismatches 23; Indels 30; Gaps 5;
 QY 2 CGRRPRVSAG-----GFADAHWTGLMTGLGEGGEGIGPEQASPTPDCAERW 50
 DB 924 CAKPREQCQCFQWVDENVAPGSAFAAPWPG---GRGAQR-----PEASKR 968
 QY 51 PRSASRPMWSAGLTVDRDRPOLGELC 75
 DB 969 PRAGSS---DAGSTVK-KPRKSLC 989

```
RESULT 12
BGAL_KLEPN
ID BGAL_KLEPN STANDARD; PRT; 1034 AA.
AC P06219;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Beta-galactosidase (EC 3.2.1.23) (Lactase).
GN LACZ.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=85289025; PubMed=3897196;
RA Buvinger W.E., Riley M.;
RT "Nucleotide sequence of Klebsiella pneumoniae lac genes.";
RL J. Bacteriol. 163:850-857(1985).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -|- SUBUNIT: Homotetramer.
CC -|- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M1441; AAA25082.1; -.
DR PIR; A24925; A24925.
DR HSP; P00722; 1BGL.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_21g.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR InterPro; IPR004200; Glyco_hydro_42C.
DR InterPro; IPR004199; Glyco_hydro_42N.
DR Pfam; PF02930; Bgal_small_C; 1.
DR Pfam; PF02929; Bgal_small_N; 1.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLYDRLASE2_N; 1.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW Hydrolase; Glycosidase.
FT ACT_SITE 468 468 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 544 544 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 544 544 117517 MW; 8F8B9DC7521EF649 CRC64;
SQ SEQUENCE 1034 AA; 14.4%; Score 65.5; DB 1; Length 1034;
Query Match 14.4%; Score 65.5; DB 1; Length 1034;
Best Local Similarity 32.7%; Pred. No. 45;
Matches 17; Conservative 6; Mismatches 20; Indels 9; Gaps 3;
QY 18 HW-----TGLTGLGEGGGIGEGGQASPTTDCASRWPSASGRWPSAGL 63
DB 983 HHKHKQAEQGWITL-DGLHMGV--GDDSWTPSVLPQWLLSQTRWQYVSL 1031

RESULT 13
CA14_MOUSE
ID CA14_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
```

```
OS Mus musculus (Mouse).
OC Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=89137932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2];
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3];
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4];
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberpaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5];
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6];
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7];
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8];
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9];
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -|- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
```

CC nidoen.

CC -!- SUSUNIT: There are six type IV collagen isoforms, alpha 1(IV) -

CC alpha 6(IV), each of which can form a triple helix structure with

CC 2 other chains to generate type IV collagen network.

CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous

CC domain (NC1) at their C-terminus, frequent interruptions of the G-

CC X-Y repeats in the long central triple-helical domain (which may

CC cause flexibility in the triple helix), and a short N-terminal

CC triple-helical 7S domain.

CC -!- PTM: Prolines are at the third position of the tripeptide repeating

CC unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which

CC are involved in inter- and intramolecular disulfide bonding. 12 of

CC these, located in the NC1 domain, are conserved in all known type

CC IV collagens.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; J03758; AAA37439.1; -

CC EMBL; M23333; AAA51625.1; -

CC EMBL; J04694; AAA50292.1; -

CC EMBL; X06777; CAA29946.1; -

CC EMBL; X02201; CAA26132.1; -

CC EMBL; M15832; AAA37340.1; -

CC EMBL; M14042; AAA37342.1; -

CC EMBL; M12879; AAA37343.1; -

CC EMBL; M13024; -; NOT ANNOTATED CDS.

CC EMBL; M13025; -; NOT ANNOTATED CDS.

CC EMBL; M13026; AAA37344.1; -

CC EMBL; M13027; AAA37345.1; -

CC EMBL; M13043; AAA37346.1; -

CC EMBL; J04448; AAA37437.1; -

CC PIR; A33525; CGMS4B.

CC MGD; MGI:88454; Col4a1.

CC GO; GO:0005604; C:Basement membrane; IDA.

CC InterPro; IPR008161; C1q helix.

CC InterPro; IPR008160; Collagen.

CC InterPro; IPR001442; Procollagn4_C.

CC Pfam; PF01413; C4; 2.

CC Pfam; PF01391; Collagen; 23.

CC ProDom; PD000007; C1q helix; 6.

CC ProDom; PD003923; ProcollagnC4; 1.

CC SMART; SM00111; C4; 2.

CC Extracellular matrix; Connective tissue; Basement membrane;

CC Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.

CC SIGNAL

FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).

FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.

FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.

FT DOMAIN 1441 1669 NONHELIICAL REGION (NC1).

FT DISULFD 1460 1551 OR 1548 (BY SIMILARITY).

FT DISULFD 1493 1548 OR 1551 (BY SIMILARITY).

FT DISULFD 1505 1511 BY SIMILARITY.

FT DISULFD 1570 1665 OR 1662 (BY SIMILARITY).

FT DISULFD 1604 1662 OR 1665 (BY SIMILARITY).

FT DISULFD 1616 1622 BY SIMILARITY.

FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 26 26 A -> P (IN REF. 2).

FT CONFLICT 186 186 S -> L (IN REF. 2).

FT CONFLICT 219 319 Q -> S (IN REF. 2).

FT CONFLICT 369 369 Q -> L (IN REF. 2).

FT CONFLICT 403 403 L -> F (IN REF. 2).

FT CONFLICT 481 481 P -> L (IN REF. 2).

FT CONFLICT 493 493 Q -> H (IN REF. 2).

FT CONFLICT 712 712 S -> I (IN REF. 2).

FT CONFLICT 813 813 E -> Q (IN REF. 2).

FT CONFLICT 982 982 Q -> H (IN REF. 2).

FT CONFLICT 1397 1397 V -> S (IN REF. 3).

SQ SEQUENCE 1669 AA; 160680 MM; 42916B91E52058E9 CRC64;

Query Match 14.1%; Score 64.5; DB 1; Length 1669;

Best Local Similarity 28.0%; Pred. No. 88;

Matches 23; Conservative 8; Mismatches 26; Indels 25; Gaps 4;

QY 2 CGPRRVVAGCGCFADAHWI-----EGQEGGIGGCGQASPTPD-- 45

Db 31 CG-----GSGCGKCDCHGVKQKGERGLGVLGVFGFMQGPGEHPGPPQKGDAGEPG 85

QY 46 -----CASRWPRASRWFSAGL 63

Db 86 LPGTGTRGPPGAGYGNPGL 107

RESULT 14

Z384_HUMAN

ID Z384_HUMAN STANDARD; PRT; 576 AA.

AC Q8TF68; O15407; Q8N938;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Zinc finger protein 384 (Nuclear matrix transcription factor 4)

DE (CAG repeat protein 1).

GN ZNF384 OR NMP4 OR CAGH1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

[1]_TaxID=9606;

RP SEQUENCE FROM N.A. (ISOFORM 1).

RA Matsuo M.Y.;

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Brain;

RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,

RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,

RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,

RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,

RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,

RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,

RA Yasuho Y., Nagai K., Isogai T.;

RT "NEDO human cDNA sequencing project.;"

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 395-576 FROM N.A.

RC TISSUE=Brain cortex;

RX MEDLINE=97369492; PubMed=9225980;

RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,

RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;

RT "cDNAs with long CAG trinucleotide repeats from human brain.;"

RL Hum. Genet. 100:114-122(1997).

CC -!- FUNCTION: Transcription factor that binds the consensus DNA

CC sequence [GC]AAAAA. Seems to bind and regulate the promoters of

CC MMPI, MMP3, MMP7 and COL1A1 (By similarity).

CC -!- SUBUNIT: Interacts with Cas (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Comment=Additional isoforms seem to exist;

CC Name=1;

CC IsoId=Q8TF68-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q8TF68-2; Sequence=VSP_006920;

CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-

CC FINGER PROTEINS.

CC -!- SIMILARITY: Contains 8 C2H2-type zinc fingers.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

DR EMBL; AB070238; BAB85125.1; -
DR EMBL; AK095734; BAC04618.1; -
DR EMBL; U80738; AAB91437.1; -
DR Genew; HGNC:11955; ZNF384.
DR InterPro; IPR007087; Znf_C2H2.
DR ProDom; PD000003; Znf_C2H2; 4.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
DNA-binding; Repeat; Alternative splicing.
FT ZN_FING 228 250 C2H2-TYPE 1.
FT ZN_FING 256 278 C2H2-TYPE 2.
FT ZN_FING 284 306 C2H2-TYPE 3.
FT ZN_FING 317 339 C2H2-TYPE 4.
FT ZN_FING 345 367 C2H2-TYPE 5.
FT ZN_FING 373 397 C2H2-TYPE 6.
FT ZN_FING 403 425 C2H2-TYPE 7.
FT ZN_FING 433 455 C2H2-TYPE 8.
FT DOMAIN 461 521 GLN-RICH.
FT DOMAIN 466 499 ALA-RICH.
FT VARSPIC 300 360 Missing (in isoform 2).
FT FTID=VSP_006920.
SQ SEQUENCE 576 AA; 63091 MW; 2A152786C3C46D90 CRC64;

Query Match 14.0%; Score 64; DB 1; Length 576;
Best Local Similarity 29.1%; Pred. No. 36;
Matches 25; Conservative 6; Mismatches 21; Indels 34; Gaps 6;

QY 11 GCGFADAHWTGLMT-----GLG-----EQEGGIGPEGQAS-----PTP----- 44
DB 41 GCGLAPHPYTLTVPASVSLPFGISMDTSKSQLTPHSQASVTQNTVVVPSTGLMT 100
QY 45 ---DCASWRP--SASRPWSAGLTV 65
DB 101 AGVSCQWRREGSQSRGP---GLVI 123

RESULT 15
TRYA_HUMAN
ID TRYA_HUMAN STANDARD; PRT; 275 AA.
AC P15157; Q9H2Y5; Q9UQL1.
DT 01-APR-1990 (Rel. 14, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-tryptase precursor (EC 3.4.21.59) (Tryptase 1).
GN TP51.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Lung;
EX MEDLINE=90009311; PubMed=2677049;
RA Miller J.S., Westin E.H., Schwartz L.B.;
RT "Cloning and characterization of complementary DNA for human
RT tryptase.";
RL J. Clin. Invest. 84:1188-1195 (1989).
RN [2]
RP REVIEWS TO 89-93 AND 108.
RA Schwartz L.B.;
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=99121069; PubMed=9920877;
```

```

RA Pallaoro M., Pejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;
RT "Characterization of genes encoding known and novel human mast cell
RT tryptases on chromosome 16p13.3.";
RL J. Biol. Chem. 274:3355-3362 (1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Lung;
RA Wang H.W., McNeill H.P., Thomas P.S., Murphy B.N., Webster M.J.,
RA Hettiaratchi A., King G., Heywood G.J., Huang C., Stevens R.L.,
RA Hunt J.E.;
RT "Molecular cloning and characterization of novel human tryptase cDNAs
RT and splicing variants.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 31-50, AND PITUITARY, SEQUENCE OF 31-38.
RC TISSUE=Lung;
RX MEDLINE=87109598; PubMed=3543004;
RA Cromlish J.A., Seidah N.G., Marcinkiewicz M., Hamelin J., Johnson D.A.,
RA Chretien M.;
RT "Human pituitary tryptase: molecular forms, NH2-terminal sequence,
RT immunocytochemical localization, and specificity with prohormone and
RT fluorogenic substrates.";
RL J. Biol. Chem. 262:1363-1373 (1987).
CC -!- FUNCTION: Tryptase is the major neutral protease present in mast
CC cells and is secreted upon the coupled activation-degranulation
CC response of this cell type.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Lys-|-, but
CC with more restricted specificity than trypsin.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Released from the secretory granules upon
CC mast cell activation.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P15157-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P15157-2; Sequence=VSP_005374;
CC -!- SIMILARITY: Belongs to peptidase family S1. Tryptase subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

EMBL; M30038; AA86934.1; -
EMBL; AF098328; AAD17846.1; -
EMBL; AF206665; AAG35695.1; -
EMBL; AF206666; AAG35696.1; -
DR HSP; P20231; IAO1.
DR MEROPS; S01.015; -
DR MEROPS; S01.143; -
DR Genew; HGNC:12018; TP51.
DR TM; 191080; -
DR GO; GO:0008236; F-serine-type peptidase activity; TAS.
DR GO; GO:0006952; P-defense response; TAS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;
KW Polymorphism; Alternative splicing.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 30 ACTIVATION PEPTIDE.
FT CHAIN 31 275 ALPHA-TRYPTASE.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY);
```

```

FT ACT SITE      224 224 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID      59 75 BY SIMILARITY.
FT DISULFID     155 230 BY SIMILARITY.
FT DISULFID     188 211 BY SIMILARITY.
FT DISULFID     220 248 BY SIMILARITY.
FT CARBOHYD     132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD     233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC      79 87 Missing (in isoform 2).
FT VARIANT       15 15 /FTid=VSP 005374.
FT VARIANT       221 221 R -> P (IN ALPHA-II).
FT VARIANT       221 221 /FTid=VAR 012102.
FT CONFLICT      215 216 K -> Q (IN ALPHA-II; dbSNP:1137382).
FT CONFLICT      215 216 /FTid=VAR 012103.
SQ SEQUENCE      275 AA; 30772 MW; B9BAC4BBCB91CE75 CRC64;

```

```

Query Match      13.9%; Score 63.5; DB 1; Length 275;
Best Local Similarity 38.5%; Pred. No. 20;
Matches 15; Conservative 3; Mismatches 8; Indels 13; Gaps 2;

```

```

Qy 30 QEGGIGPEGQASPTPDCAWRPGRGASRWPSAGLTVRDR 68
Db 28 QAGIVG--GQAP-----RSKWPQVSLRVDR 53

```

```

Search completed: August 13, 2004, 09:06:32
Job time : 15 secs

```


Result No.	Query		Length	DB	ID	Description
	Score	Match				
1	85	18.6	462	6	O97643	O97643 lama glama
2	79	17.3	813	10	O7XU80	O7XU80 oryza sativi
3	78	17.1	11096	2	Q814W3	Q814W3 streptomyc
4	77	16.9	235	4	Q8TDV3	Q8TDV3 homo sapien
5	76	16.7	1485	2	Q84IT5	Q84IT5 streptomyc
6	74	16.2	205	4	Q86SH7	Q86SH7 homo sapien
7	74	16.2	448	6	Q28936	Q28936 sus scrofa
8	73.5	16.1	311	4	Q85TK2	Q85TK2 homo sapien
9	73.5	16.1	550	10	O9LI18	O9LI18 oryza sativi
10	73	16.0	266	16	Q86639	Q86639 streptomyc
11	73	16.0	1104	3	Q9P5S3	Q9P5S3 neurospora
12	71.5	15.7	313	10	Q7XES3	Q7XES3 oryza sativi
13	71.5	15.7	597	11	Q80T82	Q80T82 mus muscul
14	71	15.6	267	16	Q82K61	Q82K61 streptomyc
15	71	15.6	359	5	Q9U3W6	Q9U3W6 drosophila
16	71	15.6	365	5	Q9W447	Q9W447 drosophila

```
DR 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSUNBA0088A01.5 protein.
GN OSUNBA0088A01.5
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Streptophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Liu Y.L., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL662987; CAD41366.1;
SQ SEQUENCE 813 AA; 90716 MW; 7B50A0B76BF8E5D4 CRC64;

Query Match 17.3%; Score 79; DB 10; Length 813;
Best Local Similarity 33.9%; Pred. No. 10;
Matches 20; Conservative 8; Mismatches 21; Indels 10; Gaps 3;

QY 3 GRPRVSGAGCGF-----ADAHWLTGEGQBG-GIGPEGQASPTPDCASRWPRGAS 55
DB 2 GDASRIDPSCGSGWRWGRGDSHHKALWSSLDGGDAEGSGPDGPA----DGAATWGRSS 56

RESULT 3
> ID Q9L4W3 PRELIMINARY; PRT; 11096 AA.
AC Q9L4W3
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nysc.
GN Nysc.
OS Streptomyces noursei.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brattaset T., Sekurova O.N., Sletta H., Ellingsen T.B., Strom A.R.,
RA Valla S., Zorchev S.B.;
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway.";
RL Chem. Biol. 7:395-403(2000).
DR EMBL; AF263912; AAF1776.1;
DR HSSP; P25715; 1MLA.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. .; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR002085; Adh zn_family.
DR InterPro; IPR000794; F:ad.
DR InterPro; IPR006162; P:plant.
DR InterPro; IPR006163; P:bind.
DR Pfam; PF00698; Acyl_transf; 6.
DR Pfam; PF00107; ADH_Zinc_N; 1.
```

```
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF0550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fabD; 6.
DR PROSITE; PS00075; ACP DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 6.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
DR PHOSPHOPANTETHEINE; Transferase.
KW SEQUENCE 11096 AA; 1150415 MW; 776CAEAFCAE551DD CRC64;

Query Match 17.1%; Score 78; DB 2; Length 11096;
Best Local Similarity 37.2%; Pred. No. 1.6e+02;
Matches 32; Conservative 7; Mismatches 33; Indels 14; Gaps 6;

QY 4 RPRVSGAGCFADAHWTGL---WTGLGEGQBGIGPEGQASPTPDCASRWPRGAS 56
DB 836 REEELSVAVTGLARAHVRGVTVRWAGLFD---GTGAREADLFTYPPQRFQREWPTAARAAQ 891
QY 57 WPWSAGLTVDRDPQLG---ELCMGRG 79
DB 892 DVTAAGLGAADHPLLGATVELADGAG 917

RESULT 4
> ID Q8TDV3 PRELIMINARY; PRT; 235 AA.
AC Q8TDV3
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative G-protein coupled receptor.
GN GPCR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Takeda S., Kadowaki S., Haga T., Takaesu H., Mitaku S.;
RT "Identification of G protein-coupled receptor genes from the human
RT genome sequence.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083587; BAB9300.1;
DR GO; GO:0004872; P:receptor activity; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000847; HTH_Lyxr.
DR PROSITE; PS00044; HTH_LYXR_FAMILY; 1.
KW Receptor.
SQ SEQUENCE 235 AA; 24154 MW; 9B1071B0D7D6B30B CRC64;

Query Match 16.9%; Score 77; DB 4; Length 235;
Best Local Similarity 37.2%; Pred. No. 4.8;
Matches 16; Conservative 3; Mismatches 18; Indels 6; Gaps 1;

QY 19 WTGLWTGLGEGQBGIGPEGQASPTPDCASRWPRGASRWPSA 61
DB 11 WLGLWVGL-----GLRFTFRVCSPLCGPLWPSASLCVWGS 47

RESULT 5
> ID Q84IT5 PRELIMINARY; PRT; 1485 AA.
AC Q84IT5
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Salinomycin polyketide synthase (Fragment).
GN SALA.
OS Streptomyces albus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1888;
```

```

RN  SEQUENCE FROM N.A.
RC  STRAIN=ATCC21838;
RA  Izumikawa M., Murata M., Tachibana K., Ebizuka Y., Fujii I.;
RT  "Cloning of Polyketide Synthase Genes Involved in Salinomycin
RT  Biosynthesis from Streptomyces albus.";
RL  Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AB087998; BAC54914.1; -.
DR  GO; GO:0016740; F:transferase activity; IEA.
DR  GO; GO:0005215; F:transporter activity; IEA.
DR  GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR  GO; GO:0008152; P:metabolism; IEA.
DR  GO; GO:0006810; P:transport; IEA.
DR  InterPro; IPR001227; AC trans.
DR  InterPro; IPR000794; Ketoacyl synth.
DR  InterPro; IPR000566; Lipocin cyteABP.
DR  InterPro; IPR006162; Peantne S.
DR  InterPro; IPR006163; Pp bind.
DR  Pfam; PF00698; Acyl transf. 1.
DR  Pfam; PF00109; Ketoacyl-synt. 1.
DR  Pfam; PF02801; ketoacyl-synt_C; 1.
DR  Pfam; PF00550; pp-binding; 1.
DR  PROSITE; PS00075; ACP DOMAIN; 1.
DR  PROSITE; PS00606; B_KETOACYL SYNTHASE; 1.
DR  PROSITE; PS00213; LIPOCALIN; 1.
DR  PROSITE; PS00012; PHOSPHOPANTHINE; 1.
FT  NON_TER 1
FT  NON_TER 1485
SQ  SEQUENCE 1485 AA; 15353 MW; 34AD09A885311B1D CRC64;

Query Match 16.7%; Score 76; DB 2; Length 1485;
Best Local Similarity 33.8%; Pred. No. 36;
Matches 22; Conservative 4; Mismatches 25; Indels 14; Gaps 3;

QY 3 GRPRVSGCGFADAHWTGLWT-----GLGEGGGGIG-PEGQASPTPCASRWPR 52
Db 1053 GRPRISRR-----SANTRGSRSRFRCPPTGTGRHRRMPGATGSPGCGFRPGR 1108

QY 53 SASRW 57
Db 1109 FEGRW 1113

RESULT 6
Q86SH7 PRELIMINARY; PRT; 205 AA.
AC Q86SH7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC046165; AAH46165.1; -.
DR EMBL; BC048278; AAH48278.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 205 AA; 22007 MW; 672587E17E2F6485 CRC64;

Query Match 16.2%; Score 74; DB 4; Length 205;
Best Local Similarity 32.1%; Pred. No. 8.5;
Matches 26; Conservative 3; Mismatches 22; Indels 30; Gaps 4;

QY 6 RRVSGCGFADAHWTGLW-TGL-----GEGGGGIGPEGQASPTPCAS 48
Db 19 RRLGAGSARVQRPAAGRWGLPGEVVVSWLEPRDGRSGRAGGGERDGLWRP-----G 73

```

```

QY 49 RWPR-----SASRWPSA 61
Db 74 RWERGARAFGEQAASRSPSA 94

RESULT 7
Q28936 PRELIMINARY; PRT; 448 AA.
ID Q28936;
AC Q28936;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Fibrinogen A-alpha-chain (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Murakawa M.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93269219; PubMed=8497848;
RA Murakawa M., Okamura T., Kamura T., Shibuya T., Harada M., Nihō Y.;
RT "Diversity of primary structures of the carboxy-terminal regions of
RT mammalian fibrinogen A-alpha-chains: Characterization of the partial
RT nucleotide and deduced amino acid sequences in five mammalian
RT species.";
RL Thromb. Haemost. 69:351-360(1993).
DR EMBL; D43760; BAA07817.1; -.
DR HSSP; P02671; IFZG.
FT NON_TER 1
FT NON_TER 448
SQ SEQUENCE 448 AA; 47355 MW; D49A3CAD1F6F0A92 CRC64;

Query Match 16.2%; Score 74; DB 6; Length 448;
Best Local Similarity 30.6%; Pred. No. 18;
Matches 26; Conservative 2; Mismatches 41; Indels 16; Gaps 4;

QY 3 GRPRVSGAG---CGFADAHWTGLWTGLGEGGGGIGPEGQASPTPCASRW-----PRS 53
Db 129 GRPEPGSTGTWDSGHDPGSGATWKP-GRPEPGSTGTWDSGHDPGSGATWKPGRPEPGS 187

QY 54 ASRW-----PWSAGLTVDPRQLG 72
Db 188 TGTWDSGRDPGSGATWKPGRPEPG 212

RESULT 8
Q86TK2 PRELIMINARY; PRT; 311 AA.
ID Q86TK2;
AC Q86TK2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to tankyrase 1 binding protein 1, 182kDa (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral Nervous System;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC046216; AAH46216.1; -.
FT NON_TER 1
SQ SEQUENCE 311 AA; 33178 MW; 4073CFAF286890F0 CRC64;

Query Match 16.1%; Score 73.5; DB 4; Length 311;
Best Local Similarity 27.6%; Pred. No. 14;

```


DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Flaggelliform silk protein-like protein.
 GN OSJNBA0073D04.40.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Euphorbiaceae; Oryzaeae; Oryza.
 OC NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA The Rice Chromosome 10 Sequencing Consortium;
 RT "In-depth view of structure, activity, and evolution of rice
 chromosome 10."
 RL Science 300:1566-1569 (2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
 RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE017097; AAP53897.1;
 SQ SEQUENCE 313 AA; 32336 MW; E749933F20DCE692 CRC64;
 Query Match 15.7%; Score 71.5; DB 10; Length 313;
 Best Local Similarity 27.5%; Pred. No. 23;
 Matches 28; Conservative 5; Mismatches 34; Indels 35; Gaps 6;
 QY 2 CGRRVRSAGCGFADAHWTG-----LWTGL-----GEGQEGGIG----- 35
 DB 26 CGRGGRPPAARPARHWRGGGGLVGEASRQRARADRGSSWPVGSRRRRRGGG 85
 QY 36 -----PEQASPTDCASRWSRSPWSAGLTVRDRPQIG 72
 DB 86 GQLSLPEAATAP-PLTRGAWPPGAS--PWRGG-SSRPRWRIG 123

RESULT 13
 Q80T82
 ID Q80T82 PRELIMINARY; PRT; 597 AA.
 AC Q80T82
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE MKIAA1858 protein (fragment).
 GN MKIAA1858
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain.
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno T., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 II. The complete nucleotide sequences of 400 mouse KIAA-homologous
 cDNAs identified by screening of terminal sequences of cDNA clones
 randomly sampled from size-fractionated libraries."
 RL DNA Res. 10:35-48 (2003).
 DR EMBL; AK122564; BAC5846.1;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004672; P:protein kinase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR000719; Prot. Kinase.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; Zf-C2H2; 3.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00028; ZINC FINGER C2H2 1; 3.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 2.

FT NON_TER 1
 FT NON_TER 597
 SQ SEQUENCE 597 AA; 64414 MW; 0D928EBD912E3951 CRC64;
 Query Match 15.7%; Score 71.5; DB 11; Length 597;
 Best Local Similarity 40.4%; Pred. No. 42;
 Matches 21; Conservative 7; Mismatches 17; Indels 7; Gaps 2;
 QY 26 LGEGQEGGIGPEQASPTDCASRWPR---SASRWP---WSAGLTVRDRPQ 70
 DB 468 LGLGHRGSGVNTGKAPLADPKSSRAPRKQATSRVPPVKSRPSGQSSRRAPQ 519

RESULT 14
 Q82K61
 ID Q82K61 PRELIMINARY; PRT; 267 AA.
 AC Q82K61
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN SAV2543.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OC NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RA "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12592562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis."
 RL Nat. Biotechnol. 21:526-531 (2003).
 DR EMBL; AP005031; BAC70254.1;
 DR InterPro; IPR008996; CytoK_IL1_like.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 267 AA; 27932 MW; 7B10A1B8389C67F8 CRC64;
 Query Match 15.6%; Score 71; DB 16; Length 267;
 Best Local Similarity 33.3%; Pred. No. 22;
 Matches 24; Conservative 7; Mismatches 27; Indels 14; Gaps 4;
 QY 7 RVSAGCGFADAHWTGLWTGLGEGQE-----GGIGPEQASPTP-----DCASRWPRG- 53
 DB 114 RVAVETGRAPA-WMLVHAMGGGQMGMSYDVGVPPQPGASAGNPVLVVRDGMPPRGR 172
 QY 54 ASRWPKSAGLTV 65
 DB 173 VVSGPQWQAVITL 184

RESULT 15
 Q9U3W6
 ID Q9U3W6 PRELIMINARY; PRT; 359 AA.
 AC Q9U3W6
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE MAB-21.
 GN MAB-2 OR CG4746.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Serano T.L., Pendleton J.D., Rubin G.M.;
 RT "A reverse genetic screen for genes involved in Drosophila
 RT development";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF214524; AAF24503.1; -.
 DR FlyBase; FBgn0029003; mab-2.
 DR InterPro; IPR004962; Mab-21.
 DR Pfam; PF03281; Mab-21; 1.
 SQ SEQUENCE 359 AA; 41220 MW; DCOB31341ACC4BS9 CRC64;

Query Match 15.6%; Score 71; DB 5; Length 359;
 Best Local Similarity 60.0%; Pred. No. 29;
 Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 39 QASPTDCASRWPRASRWP 58
 DB 169 QITPAFKCAGLWPRASRWP 189

Search completed: August 13, 2004, 09:07:24
 Job time : 43 secs